UCL

A Publication of the American Psychological Association

Journal of Abnormal and Social Psychology

UNDERGRADUATE LIBRARY

Founded by Morton Prince, 1906

CONTENTS

Editorial 461

Articles

Relation of Early Socialization Experiences to Aggression in Middle Childhood 466
Robert R. Sears

The Taking of Adult Roles in Middle Childhood 493
Eleanor E. Maccoby

Group Therapy of Phobic Disorders by Systematic Desensitization 504
Arnold A. Lazarus

Some Parameters of the Conditioning of Verbal Affect Responses in Schizophrenic Subjects 511
Kurt Salzinger and Stephanie Pisoni

Eyelid Conditioning and Three Personality Parameters 517
J. G. Field and J. C. Brengelmann

Social and Psychological Variables in Learning Hebrew 524
Moshe Anisfeld and Wallace E. Lambert

The Extinction of Fear as a Function of Distance versus Dissimilarity from the Original Conflict Situation 530
Thomas Elder, Charles D. Noblin, and Brendan A. Maher

Incubation Effect: Role of Awareness in an Immediate versus Delayed Test of Conditioned Emotionality 534
Sanford Golin

Regression and Disorders of Thought 540
Loren J. Chapman, Alvin G. Burstein, Dorothy Day, and Paul Verdone

Continued on inside front cover

This is the last issue of Volume 63.
Title Page and Contents for the Volume appear herein.

Volume 63 November 1961 Number 3

Effects of Meprobamate and of Prochlorperazine on Positive and Negative Conditioning 546
Leonard Uhr, Margaret Clay, Arthur Platz, James G. Miller, and E. Lowell Kelly
The Relationship of Persistence at a Task to Expectation of Success and Achievement Related Motives 552
N. T. Feather
Impulsive Closure as Reaction to Failure-Induced Threat 562
James E. Dittes
Motive Patterns and Risk Preferences Associated with Entrepreneurship 570
Herbert H. Meyer, William B. Walker, and George H. Litwin
Transmission of Aggression through Imitation of Aggressive Models 575
Albert Bandura, Dorothea Ross, and Sheila A. Ross
The Conditioning and Generalization of Changes in Social Perceptual Dispositions 583
Leon H. Levy
The Influence of Affective Factors on Conceptualization in Schizophrenia 588
Melvin H. Feffer
Basic Dimensions of Change in the Symptomatology of Chronic Schizophrenics 597
John E. Overall, Donald R. Gorham, and John R. Shawver
Preparatory Action and Belief in the Probable Occurrence of Future Events .. 603
Ruby B. Yaryan and Leon Festinger
Psychiatric Diagnosis: A Critique 607
Edward Zigler and Leslie Phillips
Mutual Achievement Strivings: A Continuum for Mental Health 619
Philip Lichtenberg, Rhondda K. Cassetta, and John C. Scanlon
The Influences of Personality and Situational Cautiousness on Conceptual Behavior 629
Howard A. Moss

CRITIQUE AND NOTES

Dogmatism and Acquiescent Response Set 636
Edward Lichtenstein, Robert P. Quinn, and Gerald L. Hover
The Effects of Information Source Status and Dogmatism upon Conformity Behavior 639
Robert N. Vidulich and Ivan P. Kaiman
Manifest Content of Dreams of Convicted Sex Offenders 643
Mark I. Goldhirsh
Self-Esteem, Manifest Hostility, and Expression of Hostility 646
Milton E. Rosenbaum and Robert F. Stanners
A Test of the Proposition: We Want to be Esteemed Most by Those We Esteem Most Highly 650
Cecil R. Wurster, Bernard M. Bass, and Wayne Alcock
A Cross-Cultural Investigation of the Relationship between Ethnic Prejudice, Authoritarian Ideology, and Personality 654
Aron Wolfe Siegman
Effects of Familiarity with a Controversial Issue on Acceptance of Successive Persuasive Communications 656
Edwin J. Thomas, Susan Webb, and Jean Tweedie
Anxiety and the Experimental Arousal of Affiliation Need 660
Donn Byrne
Psychogenic Deafness, Perceptual Defense, and Personality Variables in Children 663
G. F. Reed
An Outcome Index for Mental Hospital Patients 666
E. I. Burdock and Anne S. Hardesty
Status and the Judgment of Occupational Prestige 671
Sidney I. Perloe

546

552

562

570

575

583

588

597

603

607

619

629

536

539

543

546

550

554

556

560

563

566

571

Volume 63 November 1961 Number 3

Journal of Abnormal and Social Psychology

Editor: M. BREWSTER SMITH, *University of California, Berkeley*
Associate Editor: E. J. SHOBEN, JR., *Teachers College, Columbia University*

CONSULTING EDITORS

GORDON W. ALLPORT
Harvard University

URIE BRONFENBRENNER
Cornell University

ROGER BROWN
Massachusetts Institute of Technology

DONALD T. CAMPBELL
Northwestern University

LAUNOR F. CARTER
System Development Corporation

RICHARD CHRISTIE
Columbia University

MORTON DEUTSCH
Bell Telephone Laboratories

I. E. FARBER
State University of Iowa

DONALD W. FISKE
University of Chicago

N. L. GAGE
University of Illinois

ROGER W. HEYNS
University of Michigan

ROBERT R. HOLT
New York University

IRVING L. JANIS
Yale University

HAROLD H. KELLEY
University of Minnesota

LEONARD S. KOGAN
Community Service Society of New York

RICHARD S. LAZARUS
University of California, Berkeley

DANIEL J. LEVINSON
Harvard Medical School

DAVID C. MCCLELLAND
Harvard University

LEO POSTMAN
University of California, Berkeley

ELIOT H. RODNICK
University of California, Los Angeles

NEVITT SANFORD
University of California, Berkeley

SEYMOUR SARASON
Yale University

RICHARD L. SOLOMON
University of Pennsylvania

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY is devoted to both abnormal and social psychology. Its emphasis is on basic research and theory rather than the techniques and arts of practice. Abnormal psychology is broadly defined to include papers contributing to fundamental knowledge of the pathology, dynamics, and development of personality or individual behavior, including deterioration with age and disease. Articles concerned with psychodiagnostic techniques are evaluated with respect to their contribution to an understanding of the psychological principles of diagnosis, those concerned with psychotherapy are judged in terms of their contribution to an understanding of the therapeutic process. Case reports are considered in terms of their heuristic value. Reports which serve to pose or to clarify important theoretical problems, or which promise to be useful in teaching, are often published. From the social area this JOURNAL gives preference to papers contributing to basic knowledge of interpersonal relations, and of group influences on the pathology, dynamics, and development of individual behavior.

The JOURNAL is issued bimonthly in January, March, May, July, September, and November, two volumes per year. The subscription price per volume is $10.00, or $20.00 annually (foreign, $20.50). Single issues are $4.00.

Original manuscripts should be sent to the Editor-elect:

DANIEL KATZ
Department of Psychology
University of Michigan
Ann Arbor, Michigan

ARTHUR C. HOFFMAN
Production Manager

HELEN ORR
Promotion Manager

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 Sixteenth Street, N.W.
Washington 6, D. C.

Changes of address must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Mt. Royal and Guilford Aves., Baltimore 2, Md.
and 1333 Sixteenth Street N.W., Washington 6, D.C.

Second-class postage paid at Baltimore, Maryland, and at additional mailing offices

Copyright 1961 by the American Psychological Association, Inc.

Journal of Abnormal and Social Psychology
1961, Vol. 63, No. 3, 461–465

EDITORIAL

With Volume 63, my 6-year term as Editor of this journal ends, and the editorial reins pass to Daniel Katz of the University of Michigan. The wise and competent direction of the journal being thus assured for the immediate future, I report in this valedictory editorial some facts and impressions about its recent past, and record my indebtedness to the many people who have contributed to bringing the journal out. Such a report is in order, since the Editor is responsible for a "public utility," not a "private enterprise"; he holds a stewardship on behalf of the members of the American Psychological Association, of the journal's authors and readers, and less tangibly but most important, on behalf of its sector of psychological science.

To some readers, "abnormal and social" psychology has seemed to label a rather odd and ambiguous sector. When Morton Prince founded the *Journal of Abnormal Psychology* 55 years ago, he provided a vehicle for serious psychological contributions that fell outside the reigning orthodoxies then concerned with pursuit of the "generalized normal adult human mind." "Abnormal" psychology was *dynamic* psychology; it was committed to an interest in the rich if disorderly phenomena of private lives, as these come most dramatically into view from the vantage point of the consulting room. The addition of *social* psychology (in 1921) seemed natural enough. This too was a new, untidy, unorthodox field, in which the pioneer investigators, as in the case of *abnormal*, had to set greater store on the human importance of the problems than on neatness and rigor of method and theory. Joining the two fields was appropriate for an additional reason that now seems more a matter of historical accident: the earlier systematic formulations in social psychology (see the 1908 texts by McDougall and Ross) drew their explanatory principles from the same tradition of French psychopathology of which Prince was a leading American exponent. From present perspectives, the social psychology of imitation, suggestion, and crowd phenomena *was* an oddly abnormal social psychology.

Prince's main love remained abnormal psychology, however, and when he gave the journal to the American Psychological Association in 1926, the deed of gift stipulated that "the subject matter of abnormal psychology shall be preserved as the dominant feature thereof." Interpreting "abnormal psychology" broadly to include the dynamic psychology of personality processes—an extension in keeping with Prince's intent as applied to the evolving character of the field—subsequent editors have remained faithful to his charge. Social psychology has accounted over the years for less than half of the journal's contents, and editorial preference has been given to those aspects of social psychology in which focus on interpersonal or intrapersonal dispositions and processes provides linkage to the study of personality.

But times have changed, and one may well ask whether the traditional marriage of abnormal and social psychology between these green covers has the justification today that recommended it to Prince. On the sides of both abnormal and social psychology, scientific progress has transformed the fields for which the journal is responsible. Systematic investigation and explicit theory have made deep inroads on the less formal observation and leisurely speculation more characteristic of the journal a generation and more ago. By the same token, orthodoxies, especially of method, have developed on both sides of the journal's sector. Once a respectable journal of unrespectable psychology, the journal as it presently reflects its field may appear to nostalgic eyes *too* respectable through and through. And with scientific respectability and orthodoxy has gone specialization. Abnormal psychology and social psychology have lost their amateur standing. Clinical psychologists and social psychologists have developed professional identities that limit their interest in consorting with one another in a common publication.

The most insistent pressures on the marriage, practically, arise from the sheer quantitative increase in research that has accompanied these developments. Thus, in 1948, Gordon W. Allport received 149 manuscripts, of which he printed 62. I began reviewing

manuscripts in 1955, receiving 379 and accepting 146. By 1960 (when the papers published in the 1961 volumes were submitted), there were 502 manuscripts to be reviewed, of which I finally accepted 205 for publication, having reviewed 2,825 and published 990 over the period of 6 years. The results of such a growth rate virtually impose specialization and division of labor, not only in regard to editing (the task would have been impossible without Shoben's collaboration as Associate Editor in the handling of the more clinical studies), but also with respect to production and consumption of research. Authors and readers must scan ever narrower segments of literature if they are to keep abreast of it at all.

With matters like these in mind, I raised in the middle of my editorial term the question of whether some kind of formal separation between abnormal and social psychology should not be worked out within the framework of a single journal (outright divorce would be excluded by the terms of Morton Prince's deed of gift). Such a separation may well become a necessity in the not distant future, but on closer thought I am convinced that it can and should be avoided for the present. Apart from the traditions of the journal, which there are practical as well as sentimental reasons for taking seriously, several considerations pointed in the direction of maintaining the journal with its present broad spectrum of content. Foremost in my view is the desirability of resisting, in the "softer" areas of psychology with which *JASP* is concerned, the trend toward specialization along standard and narrowly compartmentalized lines.

For the scientific study of personality in its structure, development, and change cannot properly be isolated from the study of social relations and processes. The composition of this journal, as of the Division of Personality and Social Psychology of the American Psychological Association, gives institutional recognition to this fact. Grant that today's psychologists can at best master a smaller fraction of the current output that is potentially relevant to their scientific concerns than was formerly possible: but let us avoid, where we can, erecting artificial boundaries that channel institutionally this unavoidable selectivity. The lines of theoretical development in this part of our science are not clear enough, nor is the trajectory of cumulative growth in knowledge sufficiently well established, to make it a safe bet to freeze on *particular* patterns of specialization. At this choice point of editorial transition, then, *JASP* remains, wittingly, a journal of broad spectrum as it passes to the editorial care of Katz.

I have taken much satisfaction in the journal's part in publishing significant research on this wide front over the past 6 years. Nonetheless, there seems to me to be cause for concern about the health of the fields of psychology that it mirrors. The trend toward busy scientific respectability has had its costs. When I invited the Associate Editor's retrospective views, Shoben formulated the problem with respect to the more clinical aspect of the journal in words that I can only echo:

> It seems to me that a remarkably high proportion of the research reported is clean, stringently conceived, and effectively executed, reflective of rigorous and painstaking thought and experimentation, and remarkably trivial! There are occasions when I have the unpleasant fantasy that psychology has become so enamored of method that techniques become our independent variables and our substantive problems only the dependent ones. In any case, much of what has been reported in the *JASP* over the last six years indicates a domination by methods that have evolved and become elaborated along traditional lines. While there have been flashes of novelty, there have certainly been no striking breakthroughs, and I am a little appalled by the paucity of bold and vigorous theoretical thought that seeks cogency and relevance along with rigor and methodological sophistication. On the clinical side, this state of affairs implies two "disciplines" which really have too little to do with one another: One deals with traditionally researchable problems; the other with reflections on the lore of the consulting room.

Just as the greatly increased volume of submitted papers is a result of the ample financial support that has become available for psychological research, so the fixation on techniques that Shoben laments can be charged at least in part to affluence, bigness, and "projectitis." The culture of project research grants and competitive pressures from the academic and professional marketplace hardly encourages thoughtful inventiveness, for all the real advantages of the present era that we would be unwilling to forego; a spate of neat and tidy—yet conventional, fragmentary,

and not fully cumulative—reports tends to be engendered.

And as one views from a slight distance this flow of publication streaming from project grant to bibliography, one regrets a relative weakness of intrinsic direction, a deficiency in the guidance of scientific development by the growing edge of theory in the light of historical perspective. Rather, styles and fashions have tended to follow upon each other's heels in a way that reflects the "euphoria of gadgets" of which C. P. Snow says we should beware. The tachistoscope, the F Scale, the Taylor *MA* scale, the Semantic Differential: how much of the literature of the recent past revolves around a few instruments and techniques that, like Mt. Everest, had the great advantage of being *there*. Yet our psychological gadgets lack the "class" of those that furnish the laboratory of the modern physicist. Our fascination with them, the extent to which they seem to dominate our research, is hardly attributable to their intrinsic attractiveness; rather, it is symptomatic of the relative lack of clear alternative sources of direction and of the prevailing rather narrow time perspective on continuities in our discipline. Parallel observations could be made about the succession of stylish problems for research, in which shallow historical perspective has often isolated research done under a currently fashionable label from related earlier work done under different verbal auspices.

If these are indeed current liabilities of the generally prosperous fields of "abnormal and social" psychology, the journal must largely reflect them in spite of good will to buck the current where it can. During my editorial term ambitious theoretical discussion has been welcomed here—but too seldom received; continual pressure has been brought to bear on authors to present their work in the theoretical and historical context that makes its significance apparent. When authors of minor fragments were known to be engaged in substantial research programs, they were encouraged to report their work in larger segments that would warrant discussion in broader context. The Editor tried, without very much success, to convey to authors that the premium necessarily set upon *conciseness* of reporting implies no predilection for *brevity* as such. Meaty longer articles were eagerly received. A good many articles—but, in retrospect, probably not many enough—were rejected as not making a sufficiently important contribution: an uncomfortable judgment for a knowingly fallible Editor. And an increasing number of fragmentary reports that nevertheless seemed to belong in the literature were published as notes rather than as major articles. That these policies had at least some measurable effect is indicated by the fact that the mean length of major articles and case reports (notes being excluded) was 5.6 pages in 1955 but 7.3 in 1960, with the increase reflecting a steady upward trend. But the underlying problem is not to be solved by 1.7 pages.

The foregoing observations are offered not as catharsis for editorial frustrations (though they probably serve as such) nor in the presumptuous spirit I fear they may suggest; I hope rather to conserve and communicate a few of the perspectives that an Editor is bound to achieve as a by-product of hard work from his privileged vantage point, and also to make explicit the rationale for some of the policies of my outgoing regime. One final complaint: about the bad writing that stands as a barrier between so much good research and its potential reader. The fault that is at once most pervasive and most readily corrigible (by author, nor by Editor!) is that of writing that can only be characterized as *intentionally* bad; writing that skillfully follows a bad model in the mistaken notion that it is properly scientific (again the "euphoria of gadgets"!). The studiously observed and laboriously contrived passive construction; the precariously stacked multinoun phrase; the cryptic and functionless abbreviation; the "academic" stigmata of *utilize*, *it was felt that*, *more specifically*, etc.—but I need say no more.

There remains the pleasant task of acknowledging the contribution of others to editing the journal. The Editor is the focus of a large collaborative effort, to which many people commit their time and skill. I have already noted my indebtedness to the Associate Editor, E. J. Shoben, Jr., to whom I delegated primary responsibility for processing the more clinical articles. The Consulting Editors whose names appear in the front pages have devoted endless hours to reviewing manuscripts,

usually making suggestions for their considerable improvement. Most of them have performed this anonymous labor of love for a full 6 years, and there are others who served loyally as Consulting Editors during earlier years of my editorial term: Leon Festinger, Herbert H. Hyman, Gardner Lindzey, Quinn McNemar, Renato Tagiuri, and the late Harold E. Jones. I am deeply grateful to all of them.

The heavy load of reviewing manuscripts has been carried by these regularly listed Consulting Editors. But Shoben and I have called on a large number of additional persons for advice on particular papers. Listing them here not only records our gratitude for their help, but further illustrates the extent to which editing the journal has become a collective enterprise. The following persons have given us advice on one or more manuscripts:

Thelma G. Alper
Norman H. Anderson
John W. Atkinson
Albert F. Ax
Alfred L. Baldwin
Roger G. Barker
Elise S. Bartholomew
Raymond A. Bauer
Leonard Berkowitz
Jack Block
Edgar F. Borgatta
Francis J. Braceland
Jerome S. Bruner
Patricia M. Carrigan
Dorwin O. Cartwright
Joseph B. Casagrande
Dwight W. Chapman
Isidor Chein
Irvin L. Child
John A. Clausen
Lee J. Cronbach
Richard S. Crutchfield
Anthony Davids
Joel R. Davitz
Sanford M. Dornbusch
Ward D. Edwards
Charles W. Eriksen
Susan M. Ervin
Paul R. Farnsworth
Harrison G. Gough
Harry Helson
John K. Hemphill
William E. Henry
Wayne H. Holtzman
Murray Horwitz
Carl I. Hovland
J. McV. Hunt
Ray Hyman
Marie Jahoda
Jerome Kagan
Fred N. Kerlinger
George S. Klein
Otto Klineberg
Dorothy D. Lee
Louis S. Levine
Harriet B. Linton
Gerald E. McClearn
Eleanor E. Maccoby
Donald V. McGranahan
George A. Miller
Frederick Mosteller
O. Hobart Mowrer
Paul H. Mussen
Martin Orne
Charles E. Osgood
Helen Peak
Leslie Phillips
Elsa E. Robinson
Anne Roe
Milton Rokeach
Eva Rosenfeld
Julian B. Rotter
Fillmore H. Sanford
Theodore R. Sarbin
Stanley Schachter
Harold Schlosberg
Rosalea A. Schonbar
Marshall H. Segall
Claire Selltiz
Fred D. Sheffield
Edwin S. Shneidman
Robert E. Silverman
Rosedith Sitgreaves
William F. Soskin
Donald P. Spence
Herbert E. Spohn
Shirley Star
John A. Starkweather
Fred L. Strodtbeck
Donald W. Taylor
Hans-Lukas Teuber
Silvan S. Tomkins
Robert C. Tryon
Hans Wallach
Robert W. White
Robin M. Williams
Frederick Wyatt
Gertrud L. Wyatt
Robert B. Zajonc
Joseph Zubin

The arduous task of preparing copy for the printer and checking proof has been ably performed in the Central Office of the American Psychological Association under the direction of the Managing Editor of the association's publications: initially, Lorraine Bouthilet, and during the last 5 years, Arthur C. Hoffman. To my competent secretaries at New York University and at Berkeley, Susi Stein and Sara F. Caldwell, I am indebted for handling the journal's extensive correspondence, maintaining its complex files, and, on occasion, sustaining my editorial morale.

It is probably inevitable that the Editor will often have the feeling of flying blind: processing a monstrous mail and shipping stacks of manuscripts to the APA publication office whence they eventually emerge in print, with only an occasional critical howl or appreciative comment. The "feedback" that does come to him, from authors and readers, whether positive or negative, is therefore particularly welcome, and I should like to record my special gratitude to those authors who have not only submitted gracefully to what must have often appeared to them as editorial mayhem, but have taken the trouble to let me know that in retrospect they thought their intellectual offspring may have benefited

from the harsh treatment. Write to the new Editor when the spirit moves you!

The pressures, the frustrations and gratifications of the editorial role predictably leave me ambivalent as I turn over the direction of the journal to Daniel Katz. But there is no ambivalence in my attachment to the venerable journal, and it is therefore very satisfying to know that the new Editor is so admirably qualified to guide it toward a distinguished contribution to psychology in the coming years. May his editorial regime prosper, and may abnormal and social psychology flourish!

M. BREWSTER SMITH

Journal of Abnormal and Social Psychology
1961, Vol. 63, No. 3, 466-492

RELATION OF EARLY SOCIALIZATION EXPERIENCES TO AGGRESSION IN MIDDLE CHILDHOOD[1]

ROBERT R. SEARS[2]
Stanford University

This report presents one part of a follow-up study of the 5-year-old children whose early socialization experiences were described in *Patterns of Child Rearing* (Sears, Maccoby, & Levin, 1957). It attempts to trace the effects of some of those experiences on the children's aggressive behavior 6 years later.

During the winter of 1951–52, 379 public school kindergarten children in Newton and Watertown, Massachusetts, were selected as subjects for a study of the child rearing antecedents of identification. The design of that research required the measurement of certain aspects of both the children's behavior and the mothers' child rearing practices and attitudes. The measures of identification which seemed appropriate, at that time, related chiefly to the extent of children's adoption of adult role behavior and to the stage of development of guilt feelings and other signs of conscience. Measures of such behavior were obtained from three sessions of doll play, two permissive and one structured (Levin & Sears, 1956; Sears, 1953), and from mother interviews (Sears et al., 1957). These interviews were mainly directed to the measurement of the antecedent (maternal) behavior, however; they were designed to secure information about aspects of the mother's practices and attitudes that had theoretical relevance to the development of identification. The dimensions measured related to early caretaking, feeding, toileting, the handling of dependent, aggressive, and sexual behavior, and methods of discipline. These interviews were based on a standard series of questions and probes, and permitted entirely free answers. They were secured by trained interviewers in the mothers' own homes, were recorded and transcribed, and then coded on 188 scales. It is the mothers' scores on these scales that provide the antecedent measures in the present longitudinal study of the relation of early child rearing experiences to children's manifestations of aggression at age 12.

Follow-Up Study

While the original study was directed to the problem of identification, the theoretical position on which the research was based required measurement of maternal behaviors that were equally relevant to the development of various other aspects of children's personalities. Since the extent and kind of influence of early socialization experiences on later childhood and adult behavior have been studied mainly by clinical techniques, this body of data seemed of considerable potential value both for testing hypotheses concerning developmental relationships and for searching out new ones. What was needed was a set of appropriate measures of the children, at age 12, which could be compared with these earlier obtained reports of their socialization experiences. To these ends, therefore, a follow-up study was planned that would provide measures not only of such presumably identification-induced behavior as adult role choice, resistance to temptation, and sex typing (masculinity-femininity), but of aggression and dependency also.

METHOD

Procedure. The follow-up measures of the children were obtained in the spring of 1958, when the children were in the sixth grade and were mainly between 11-6

[1] From the Laboratories of Human Development of Stanford University and of Harvard University. This is the first of several reports on a follow-up study of the children described in *Patterns of Child Rearing* (Sears, Maccoby, & Levin, 1957).

The substance of these findings was presented in the first John E. Anderson Lecture at the Institute of Child Development and Welfare of the University of Minnesota on March 7, 1961.

[2] The research was designed in collaboration with Eleanor E. Maccoby, John W. M. Whiting, and Wesley Allinsmith. Eleanor Maccoby and Wesley Allinsmith supervised the collection of the final data with the assistance of Roger V. Burton, and Eleanor Maccoby has collaborated in its analysis. I wish to express appreciation also to John C. Wright, who, while at Stanford, assisted in the preparation of the aggression tests. The earlier stages of this research, at Harvard, were supported largely by a USPHS grant (M 461). The follow-up study was supported by an additional grant to Harvard (M 1468) and a mental health program grant to Stanford University by the Ford Foundation.

and 12-6 years of age; for communicative convenience they will be called "12-year-olds" (cf. Table 1). This year was chosen because the schools operated on a 6-3-3 division, and the sixth grade was the last occasion on which the children would be together in a single-teacher grade, permitting ratings of school behavior by a teacher who had full time supervision of a given child.

The measures obtained were of four kinds. First, there was a series of self-administering scales, constructed or adapted for the present purpose, for the measurement of masculinity-femininity, self-concept, nurturance, adult versus child role choice, and five aspects of aggressive attitude. Second, standardized teacher interviews were recorded, transcribed, and coded for the measurement of other displays of a few of these same variables. Third, a story completion measure of guilt was obtained. And fourth, a realistic test situation for the measurement of resistance to temptation was administered (Grinder, 1960). The present report describes only those findings relating to aggression which were obtained from the self-administered scales. Although an attempt was made to include measures of the children's aggression in the teacher interviews, the teachers were unable to report on this aspect of behavior with sufficient clarity to provide useful data. This may have resulted, in part, from the institutional nonpermissiveness for child aggression toward either teacher or peers at the sixth grade level. The major difficulty in securing teacher interview measures stemmed from the lack of discriminable differences.

Subjects. Of the original 379 children studied (Sears et al., 1957), approximately 200 had moved out of the school districts in which they had originally lived; to arrange for securing their cooperation and that of the schools and teachers in their new locations promised a very heavy expense. Therefore, the new measures were secured from only those children who were enrolled in the sixth grade classrooms of the schools, in the original towns, in which seven or more of the sample were currently enrolled. The consequence of this selective procedure was that completed aggression attitude scales were obtained from 160 of the original 379 subjects, and also from 377 other sixth graders who had not been in the original study. In the following discussion, the 160 "originals" who had been subjects (Sears et al., 1957) are referred to as the *follow-up sample*, and the others as the 1958 *classmates*. The teachers were interviewed only about members of the sample group—and of course only for that group are the child rearing measures from mother interviews available—so that data analyses involving internal comparisons within the self-administered measures will be presented for the *combined* groups, follow-up sample and classmates, but analyses involving antecedent-consequent comparisons from the 1951–52 mother interviews must be limited to the follow-up sample group.

Results

Since the drop from 379 to 160 cases is a large one, some importance must be attached to questions concerning the sample composition. In Table 1 are summarized certain demographic properties for three groups: the original group, the follow-up sample, and the 1958 classmates in the sixth grade. Reports of ordinal position and father's occupation had been secured from the mothers of the original group, in 1951–52, and were obtained from the children themselves in 1958. Hence double figures can be presented for the follow-up sample, those obtained at kindergarten age,

TABLE 1

SOME DEMOGRAPHIC CHARACTERISTICS OF THE ORIGINAL (1951–52) AND THE FOLLOW-UP (1958) POPULATIONS
(In percentages)

Characteristic	1951–52 Original group		1958 Follow-up sample				1958 Classmates	
	1952 measures		1952 measures		1958 measures		1958 measures	
	Boys (N = 202)	Girls (N = 177)	Boys (N = 76)	Girls (N = 84)	Boys (N = 76)	Girls (N = 84)	Boys (N = 189)	Girls (N = 188)
Ordinal position:								
Only	16.3	10.7	21.0	6.0	18.7	3.6	6.7	4.6
Oldest	29.2	28.3	26.3	25.0	30.7	28.6	31.8	36.5
Middle	23.8	32.2	21.0	35.7	24.0	40.4	21.8	30.3
Youngest	30.7	28.8	31.7	33.3	26.6	27.4	39.7	28.6
Father's occupational status[a]:								
Upper middle	65.2	56.9	60.0	58.5	63.4	56.6	53.5	57.3
Lower middle	8.6	10.3	14.7	11.0	14.1	9.6	15.1	10.5
Upper lower	26.2	32.8	25.3	30.5	22.5	33.8	31.4	32.2
Age at 1958 testing:								
M					11–8	11–10	11–11	11–10
SD (months)					4.6	3.0	7.0	4.8

[a] Father's occupational status was measured on the Warner (1952) scale and on the Hollingshead (1958) scale. Both are seven-point scales. Upper middle includes Categories 1–3, Lower middle is Category 4, and Upper lower includes Categories 5–7.

for comparison with the total original group, and those obtained in the sixth grade, for comparison with the 1958 classmates groups. In comparison with the total original group, the follow-up sample contained a slightly greater proportion of the original "only" boys, a slightly lesser proportion of the "only" girls, and slightly greater proportion of the "youngest" girls. None of these differences seems of major import, particularly in light of the fact that a few more children were born to these families during the intervening years.

It is interesting to compare the relatively unselected group of sixth grade classmates with the sample. As of 1958, the sample was biased in favor of more "only" and fewer "youngest" boys, and more "middle" and fewer "oldest" girls. This bias reflects the effort made in the original study to select a stratified sample providing a more nearly equal composition in terms of sex, ordinal position, and socioeconomic status (cf. Sears et al., 1957, pp. 24–27, 512).

With respect to social status, comparisons have been made on the basis of father's occupation. The Warner Classification (Warner, Meeker, & Eells, 1949) was used in the original study, the Hollingshead (Hollingshead & Redlich, 1958) for the follow-up. The two are sufficiently similar for the present comparison,[3] which shows that the follow-up sample is closely representative of the total original group. However, comparison of the 1958 sample and classmates boys shows that the sample population contained a moderately (10–11%) greater representation of upper middle class children than the classmates. Obviously the original group did, too. There is no difference between sample and classmates girls, and both are closely representative of the original 1952 group. Mean ages of the two 1958 groups of boys differed by 3 months, the sample being younger ($p < .001$)[4]; the girls' groups were identical. Both the sample groups were narrower in range, however, as would necessarily be the case; they had all been in kindergarten in 1952 and were all in the sixth grade in 1958. Children who became retarded or accelerated during the 6-year interval were not reached by our follow-up procedure. The unselected sixth grade classmates, on the other hand, contained both retarded and accelerated children, and hence the range of ages was greater.

So far as these various demographic measures are concerned, then, the follow-up sample is a surprisingly close representation of the total original group from which it was drawn. Since none of these three populations can be considered accurately representative of any wider population, the comparability among them (except for the heavier representation among upper middle class boys) permits us to turn with not too much worry to within-group comparisons.

Here we find further reassurance of the relatively unbiased character of the follow-up sample. Comparison was made of the means and distributions of 22 mother interview scales for the 160 cases in the follow-up sample and the 219 cases not followed-up. Six of these scales described the child's behavior (amount of weaning upset, severity of feeding problems, bed-wetting, amount of aggressive behavior in the home, amount of current dependency at kindergarten age, amount of conscience); none of these showed any difference. The other 16 scales related to important maternal behaviors. On only two of them was there a significant difference between the mean ratings; toilet training was a little less severe ($p < .02$) and the use of deprivation of privileges was less frequent ($p < .03$) in the follow-up sample. There were no differences in early feeding, the handling of sex, aggression or dependency, the use of other disciplinary methods, or in family atmosphere, authority, and warmth. There was no difference, either, in the amount of aggression the children had shown on the first session of doll play in 1952.

In sum, the follow-up sample was a pretty accurate representation, both demographically and with respect to measures of child behavior and child rearing, of the original group whose socialization experiences were reported in *Patterns of Child Rearing* (Sears et al., 1957). Further, with the exception of a 10% over-

[3] The Warner index of SES, which was used in the original study (Maccoby & Gibbs, 1954), could not be used for the follow-up because it included income as one variable. Since we lacked the resources for further inquiry from the mothers, we contented ourselves with father's occupation, which the children could report. Occupation is weighted two and income, one, on the Warner revised index of SES.

[4] Unless otherwise indicated, all p values are based on a two-tailed test.

weighting of upper middle class boys, and three variations of 10–12% each in ordinal position weighting, the sample was very similar demographically to the large group of their sixth grade classmates.

MEASUREMENT OF AGGRESSION

How does one measure aggression in 12-year-olds? Clearly, the answer rests on what theory is adopted as to the genesis and dynamics of such behavior. There have been two main views about the matter that have influenced measurement and interpretation in both the clinical and experimental settings. One of these is instinctual and drive oriented; the other rests on a general theory of action and learning. Neither has been spelled out in sufficient detail to permit a purely theoretical derivation of appropriate methods of measurement. However, researchers who lean toward the instinctual view have tended to emphasize the use of projective methods, while those who find action theory more congenial have tended to use social behavioral measures. A brief comment on each will make clear the reasoning that led to the choice of instruments in the present study.

The notion that there is a certain quantum of internal instigation to aggression has been inherent in much of psychological theory for some time. It was expressed most explicitly first by Freud (1922) in his conception of the death instinct. This nativistic position gave little attention to the origins of individual differences in amount or quality of aggressive behavior except by implication; childhood experience and environmental influences could excite or inhibit the expression of the instinctual impulse, and could displace it outward. From the measurement standpoint, such a theory seems to imply that there are various alternative behavioral indices of the underlying motive. The chief aim in choosing one rather than another is to obtain the freest (least inhibited) expression. Thus, a favorite device for the measure of many types of primary or secondary "drive induced" behavior is some projective technique such as the TAT, story completion, or doll play. These methods are assumed to aid in circumventing social inhibitions that would otherwise suppress the full expression of the drive. They represent an effort to sample the "true" uninhibited quantum. However, while there is no question that projective responses do reveal the existence of instigation to aggressive, sexual, achievement, and other socially controlled actions, there is also good evidence that such projections are themselves strongly influenced by momentary conditions of excitation and inhibition (Clark, 1955; R. R. Sears, 1951).

The alternative theory makes no assumption about a quantum of internal instigation that must be uninhibitedly tapped. On the contrary, it presumes that whole classes of aggressive acts are learned as responses to many different stimulus situations. By stimulus generalization, these acts come to be evoked by great ranges of stimuli, both internal and external, including ones that are symbolic or have imaginal qualities only. By response generalization, acts can shift from one action category to another (e.g., from physical to verbal) or from one psychological function to another (e.g., from overt action to imagination or fantasy). Indeed, the lability of expression is so great that even entirely novel stimulational settings and manipulanda seem able to elicit and provide a means of expression of aggressive acts. Each such setting has its own instigative value, deriving from direct reinforcements in the past, as well as from primary or secondary stimulus generalization. This effective instigative value is a cumulation of both excitatory and inhibitory values, of course.[5]

This emphasis on stimulus situations need not exclude internal instigators, including those having essentially the qualities of drives. Indeed, in one respect, *Frustration and Aggression* (Dollard, Doob, Miller, Mowrer, & Sears, 1939), which first formally presented the action and learning theory of aggression, was not at all dissimilar to the earlier instinctual view; it simply relocated the source of impulsion from instincts to a frustration drive created by the cumulated frustrations of motivated behaviors of all kinds. The internal source was changed from a native reservoir of

[5] McDougall's (1908) instinct of pugnacity rested somewhere between these two views. It did not assume a pool or reservoir of internal instigation (drive), but neither did it rely explicitly on learning principles to account for individual differences in directions and qualities of aggressive acts; these were instinctually organized action patterns.

energy to an experience-induced drive state. More importantly, however, this view introduced individual differences into the source of internal instigation, and placed the internal instigator in the same category of operationally definable antecedents as the external ones.

The relevance of these theoretical matters to the problem of measurement lies in the assumptions that must be made about what is being measured. From a drive theory, one seeks to tap that way of behavioral expression of the drive that suffers least interference from either internal controls (e.g., anxiety, conscience) or external threateners (e.g., authority or moral symbols). The acceptability of any index can be evaluated either by a priori analysis of the measurement situation with respect to freedom from inhibition, or by appeal to some criterion against which the index can be validated. Projective techniques commonly have received acceptance on a priori grounds—mainly, it is to be feared, as a result of uncritical analysis of the inhibiting potentialities of the test situation—or their validity has been tested against other behavioral measures which are themselves accepted on a priori grounds.

With an action and learning theory, the measurement problem is somewhat different. Measurement in this case means discovering the strength of specifically defined aggressive actions in specifically defined stimulating conditions. It is assumed that the measured "strength" or "amount" is a product of all past direct reinforcements and inhibitions plus all relevant stimulus and response generalizations of both excitation and inhibition. A measure is evaluated by three criteria: agreement by theorists that the measure is relevant to their interests and is not grossly mislabeled, provision of a consistent response value under standard conditions (i.e., adequate reliability), and lawful relationships to independently measured antecedent learning conditions or coincident stimulating conditions. Most measures of aggression, both projective and otherwise, meet the first criterion—if the theorists collected for judging are not too heterogeneous a lot. Reliability is more difficult. Projective methods, particularly those that are susceptible to the influence of permissiveness in their administration, do not usually provide high coefficients on a test-retest basis, though odd-even reliability may be satisfactory. But the most important criterion is the third. If stable and theoretically meaningful relationships can be found between a given measure of aggression and some past experience or concurrent situation, the measure itself must have adequate reliability even though this cannot be demonstrated by formal methods of evaluation. Single acts that cannot be repeated (e.g., a lynching), and acts the potentiality for which suffers modification simply as a result of their own performance (e.g., doll play aggression), provide special difficulty in this respect. Quite a number of them have nevertheless provided useful antecedent-consequent relationships.

To return, now, to the question of how to measure aggressiveness in sixth grade children. There are many types of aggressive actions and many instigating situations that evoke them. The majority of these acts are impossible to measure directly; overt aggression toward mother in the absence of the father, self-aggression when alone, spontaneous fantasy aggression toward symbolic authority figures, etc., do not lend themselves readily to measurement, reliable or otherwise, even though they might be excellent choices from the standpoint of the first and third criteria. What does seem practical, however, is the use of standardized forms of self-report of attitudes, and the securing of reports of overt behavior from outside observers. For the latter, there are three main sources that would be useful to have—parents, teacher, and peers—because of their systematically different stimulus significance to a child. The procedural limitations of the present study led us to seek only the teacher observations, and they proved useless with respect to aggression. For self-reports of attitudes, scales were constructed to measure five aspects of aggression, and they proved useful.

Aggression Scales

These scales were composed of items in the form of declarative sentences with which the subject could express his agreement or disagreement on a five-point scale ranging from "strongly agree" through "not sure" to "strongly disagree." The content of each item was designed to express one type of aggression (or its rejection). The presumption was that

the reading of a statement would arouse an imaginal representation of the act or feeling described, and that marking the scale would be, by response generalization, an affirmation or rejection of that representation. A score of 5 was given to the "strongly agree" or "strongly disagree" end of the five-point scale that represented acceptance of the type of aggressive action or feeling portrayed by that item. The next position received a 4, and so on. A total score for each of the five scales was obtained by cumulating these individual item scores for all the items belonging to a given aggression scale. The five scales were as follows:

Aggression anxiety. This scale measured feelings of fear, discomfort, and dislike of aggression. The aggressive stimulus to the arousal of such feelings was placed, in some items, in the person himself ("It upsets me to think some thoughtless word or crack of mine might hurt someone's feelings") and sometimes outside the self ("It makes me uncomfortable to see two of my friends fighting"). These external stimulus conditions ranged from the personal or social, as in the last example, to relatively impersonal or symbolic external events ("If someone gets hurt in an auto accident, I usually try to get a good view of what happened"). The signs of the anxiety in response to aggressive stimulation included references to avoidance behavior, somatic symptoms, guilty feelings, expectations of injury or trouble, and the use of descriptive words indicative of strong negative feelings ("vicious crimes").

Projected aggression. This scale measured the tendency to attribute aggression to sources outside the self. The items included a range of *objects* of hostility, including the self. The *agents* of aggression ranged from specific ("Big dogs are likely to be dangerous") to very general ("Sometimes I feel that there are so many bad people in the world that even my own town is about as dangerous to live in as a real jungle"). The agents to whom aggression was attributed included people, animals, and natural forces. About half the items involved aggression toward the self from other sources, and the other half reflected aggression toward other objects. Of those which reflected aggression toward the self, some involved derogation, and some represented overt aggression.

Self-aggression. All items referred to injury or punishment to the self. Two described impulses toward suicide, two related to self-punishment from other sources than the self, and one referred to accidents to the self.

Prosocial aggression. The dimensions represented by this scale and the next one were first described by Johnson (1951) in her study of doll play aggression in 5- and 8-year-old children. Prosocial aggression is aggression used in a socially approved way for purposes that are acceptable to the moral standards of the group. The items referred to law enforcement, strictness of control, punishment for rule breaking, and insistence on appropriate rules about aggression ("When a person has broken an important rule, he should definitely be punished for it"). A high score on this scale is a positively aggressive score, but the form of aggression is in favor of socially acceptable controls and disciplines. In its extreme form, prosocial aggression probably represents "moral righteousness." However, a low score could be the product of either one of two characteristics in the person: a low level of aggressive reactivity, or a tendency to be aggressive in unacceptable ways, with resistance to authority or the rules of the group. Four of the eight items that composed the scale would be influenced by this rejection of authority, and four would not be. In the absence of other information, then, a low score on this scale cannot be clearly interpreted. A high score does clearly mean a high aggressiveness of this socially accepted sort.

Antisocial aggression. This dimension contrasts with the previous one in the sense that the aggressions referred to here are ones that are normally unacceptable socially in the formal pattern of our culture. Some items included a reference to the positive desirability of powerful aggressive motivation, others to the acceptability of ordinarily disapproved forms of aggression, and still others to the "naturalness" of aggression as a form of social interaction ("Sometimes an actual fight is the only way to settle an argument"). In terms of our contemporary middle class American culture, antisocial aggression represents a much lower level of socialization than does prosocial

aggression. A high score may be interpreted as indicating a strongly aggressive person without social control. A low score may describe a person without much aggressive reactivity, or one whose aggression takes other forms, such as the prosocial.

Scale Construction and Item Validation

Seventy-three items were prepared for initial testing. These were divided unevenly among six scales,[6] the number ranging from 6 for self-aggression to 18 each for aggression anxiety and projected aggression. Twenty buffer items were constructed also; these related to such nonaggressive matters as school activities, career choices, eating, etc.

There were two item validation series, both performed with sixth grade classes in two school districts in the San Francisco Bay area. These classes were about equally composed of boys and girls, and had a wide socioeconomic range. For the first testing, the items from the first three scales (plus the ultimately rejected "attenuation" items) were combined with the buffer items, in an alternation order, and presented to two classes containing 58 children; the last two scales were handled in the same way, using two other classes of the same size.

A priori scoring of the items provided a total score on each scale for each child tested with that scale. Odd-even reliabilities were low but promising. A rough item validation was performed by dividing the groups in half on the basis of total score for each scale and counting the frequency with which each item was scored in the appropriate direction by the upper and lower half of the group. Several items proved to be nondiscriminating at the .10 level and were discarded in preparing the test forms for the second validation series.

For the second series a substantial change was made in the form of the test blank. The retained items for the aggression scales were combined with the 20 buffer items, which were introduced among them in such a way as to prevent too severe piling up of negative emotions by the repeated reference to aggression. The items for the various aggression scales were distributed among one another in order to avoid establishing response sets on any one dimension. Attention was also given to randomizing positively toned and negatively toned items for which the high aggression answer would be "disagree" and "agree," respectively.

This test blank, containing a total of 80 items (including the 12 ill-fated attenuated aggression items), was presented to two other classes of sixth grade children with a total N of 57. Odd-even reliabilities were substantially improved. Item validation was again performed, using for each scale, separately, the upper and lower half of the group on total score. Items that failed to be discriminative at the .05 level were discarded for the final form of the test. Reliability coefficients for the final form are reported in Table 2.

Collection of Data

For the measures taken in Newton and Watertown, the test blanks were lithographed in the same format as had been used for the second validation series.[7] Various test instruments for the follow-up study were given to whole classes at a time on separate days. The aggression scales were the third such presentation. The first two had also been self-administering scales; they had been presented by research assistants who were introduced by the teachers as being from Harvard. Complete confidentiality of the test blanks had been assured, with emphasis that the teacher would not see them. For the aggression scales, the presentation was made by the teacher but with similar emphasis on confidentiality. The children placed their own blanks in an envelope which was then sealed and taken to the principal's office.

[6] An effort was made to construct a scale for "attenuated aggression," but this conceptual dimension proved to have inadequate construct validity and the scale had poor internal consistency. The effort was discontinued, therefore, though the items of the hoped-for scale were retained in the test blank in order to secure data which could be analyzed for other purposes.

[7] A keyed copy of the test blank, together with a table showing proportion of each sex in the 1958 Classmates group who scored at each of the five points on each of the 60 items, and a complete table of inter-item correlations calculated separately by sex has been deposited with the American Documentation Institute. Order Document No. 6870 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $2.25 for microfilm or $5.00 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

TABLE 2
ODD-EVEN RELIABILITIES OF THE FIVE AGGRESSION SCALES: 1958 CLASSMATES
(Sexes combined)

Scale	Number of items	Number of cases	Pearson r	
			Uncorrected	Corrected
Aggression anxiety	12	363	.44	.61
Projected aggression	14	361	.44	.61
Self-aggression	5	348	.08	.15
Prosocial aggression	8	364	.46	.63
Antisocial aggression	9	364	.49	.64

Reliability

The 1958 Classmates (i.e., the sixth graders who participated in this testing but were *not* members of the follow-up sample) have been used as a group on which to determine the odd-even reliabilities of the five scales. In Table 2 are shown the numbers of items remaining in each scale after the second validation series, the number of children on whose papers reliability scores could be calculated, and both the uncorrected and corrected (Spearman-Brown) coefficients of reliability.

The corrected reliability coefficients for aggression anxiety, and projected, prosocial, and antisocial aggression are all .6 or better. That for self-aggression is significant at just the $p = .01$ level, which is scarcely promising for a reliability coefficient! Furthermore, the 10 correlations among the five items, for the 174 classmates boys, range from −.12 to +.31, with a median of +.02; for the 170 girls, the range is −.14 to +.28, with a median of +.12. Nevertheless, as will be seen later, the boys' scores on this scale appear to have certain theoretically predictable relationships with child rearing antecedents, and it may be that, as in the case of certain aptitude tests, the total score has greater stability than is implied by the intercorrelations among its components. Obviously, later results on self-aggression must be interpreted with considerable caution.

RESULTS AND DISCUSSION

Intercorrelations among Scales

In Table 3 are shown, separately by sex, the intercorrelations among the five scales for the total 1958 group, both follow-up and classmates combined. In Table 4 are shown, for the follow-up sample, the intercorrelations among not only the aggression scales but also four measures of child behavior obtained from the mother interviews 7 years earlier. The intercorrelations among the five aggression scales are similar for the follow-up sample and the total population except in one instance; boys' projected and antisocial aggression are correlated .23 in the follow-up sample but only .04 in the total population (which includes the sample). The theoretical implications of these matrices will be discussed in connection with results of the antecedent-consequent comparisons.

TABLE 3
INTERCORRELATIONS OF AGGRESSION SCALES: TOTAL TESTED POPULATION
(Girls: $N = 264$; Boys: $N = 261$)

Boys \ Girls	Scale	1	2	3	4	5
Aggression anxiety	1		.19	.08	.45	−.44
Projected aggression	2	.35		.44	.18	.17
Self-aggression	3	.24	.39		−.02	.07
Prosocial aggression	4	.50	.13	.12		−.15
Antisocial aggression	5	−.39	.04	.02	−.32	

Note.—Girls' figures above the diagonal, boys' below.

Sex Differences

In Table 5 are shown, for the combined groups, the means and *SD*s of the scores on the five aggression scales for each sex, together with the significance of the sex differences. The girls are significantly higher on prosocial aggression and aggression anxiety, while the boys are substantially higher on antisocial aggression. There are no sex differences on the other two scales.

As was mentioned earlier, a test of masculinity-femininity was also administered to the children. It was constructed from the 38 items of the Fe scale in Gough's CPI (1957). That scale was revised and revalidated for the present study to make it more suitable for this younger age group. The scale as used had quite satisfactory odd-even reliability, and intrasex validity of the 33 individual items retained from the original form was uniformly acceptable, as was the intersex validity of all but four items.

In the present connection, it is interesting to note that the correlations within each sex

TABLE 4
INTERCORRELATIONS OF AGGRESSION SCALES AND FOUR AGE 5 MEASURES OF CHILD BEHAVIOR: FOLLOW-UP SAMPLE
(Girls: $N = 84$; Boys: $N = 76$)

Boys \ Girls	Scale	1	2	3	4	5	6	7	8	9
Aggression anxiety	1		.19	.11	.40	−.48	.18	.04	−.24	.00
Projected aggression	2	.23		.38	.14	.23	−.06	−.03	−.16	−.05
Self-aggression	3	.25	.42		−.01	−.01	.04	.05	−.07	.24
Prosocial aggression	4	.61	.05	.10		−.03	.04	−.08	−.40	.18
Antisocial aggression	5	−.33	.23	.14	−.36		−.08	−.10	−.03	−.12
Aggression: age 5	6	−.21	.05	−.10	−.18	−.03		.14	−.14	.02
Dependency: age 5	7	−.16	.01	.03	.00	.06	.02		.02	.17
Conscience: age 5	8	.30	.08	.00	.04	−.06	−.16	.06		.02
Bed-wetting: age 5	9	−.17	.05	−.18	−.03	.00	.10	.10	−.37	

Note.—Girls' figures above the diagonal, boys' below.

between femininity and both prosocial aggression and aggression anxiety are significantly positive, and those between femininity and antisocial aggression are significantly negative. Thus, the directions of differences in means *between* sexes are similarly reflected in directions of correlations *within* each sex.

These findings suggest that high prosocial aggression is a feminine quality, and high antisocial aggression is a masculine one. Coupled with the feminine reluctance to express this latter type of attitude is a higher level of aggression anxiety.

Comment. These sex differences in kinds of aggressive feelings are an objectification of a rather common cultural stereotype. Girls are more fearful, more passive, more conforming to social norms and proprieties; boys are more violent, rougher, and more destructive. It is impossible to tell to what extent our measuring instrument reflects an expression of inner feelings and to what extent it provides opportunity for the acquiescent expression of socially appropriate attitudes. Doubtless both factors influence the responses children make in such a situation. On the one hand, as our theory of measurement suggests, the questionnaire items probably evoke imagery and feelings that are stable and relatively long-enduring; but on the other, one must recognize that the very setting in which the scales are presented contains stimuli appropriate for eliciting stereotyped reactions. The children are seated among their peers of both sexes; they are in the school room; they are faced and instructed by adults. Whatever attitudinal expressions have become attached to these stimulus conditions, whether "stereotypes" or not, are elicited in some degree.

TABLE 5
SEX DIFFERENCES ON THE FIVE AGGRESSION SCALES: TOTAL TESTED POPULATION
(Girls: $N = 264$; Boys: $N = 261$)

Scale	Girls		Boys		Sex differences	CR
	M	*SD*	*M*	*SD*		
Aggression anxiety	6.11	1.49	4.82	1.63	1.29	9.46
Projected aggression	4.51	1.87	4.53	1.90	.02	.123
Self-aggression	4.13	1.65	4.27	1.61	.14	.986
Prosocial aggression	5.46	1.48	4.81	1.74	.65	4.61
Antisocial aggression	4.75	1.66	6.09	1.82	1.34	8.80

Note.—Figures given here are based on a scale of 1-9 that covered the full range of obtained scores on each test.

This distinction between feelings and stereotypic attitudes may or may not be important for an understanding of the structure of aggression. But since our present main interest is in the effects of early experience on these behaviors, the issue can be held in abeyance while we discover whether there are, in fact, predictable (or detectable) relations between these measures and the hypothesized antecedents. If so, then the question that needs to be raised—as with any measure—is as to how widely these particular measures may be generalized to other instances of aggressive behavior. The remainder of the present study, however, is devoted to the antecedent-consequent problem.

Permissiveness and Punitiveness

The effect of maternal punitiveness on preschool children's aggressive behavior has

been shown in two previous studies. In the earlier (Sears, Whiting, Nowlis, & Sears, 1953), the mothers of 40 3- and 4-year-olds were interviewed with respect to their control of aggression. Ratings of the severity of their customary punishments were made from transcribed interviews. The amount of aggressive behavior shown by the children in a preschool was measured both by teacher ratings and by direct observation. For boys, the relationship between amount of aggression shown and the severity of maternal punitiveness was clearly positive. For girls, however, the relationship was curvilinear, both the most and least severely punished girls being less aggressive than those suffering an intermediate amount of punishment.

An hypothesis was suggested that punishment serves as a form of frustration and hence increases the total instigation to aggression, but that when punishment becomes sufficiently severe, it inhibits the specific actions punished; in such cases, the increased aggressive instigation would be manifested only in forms of aggressive activity different enough from those punished not to suffer from inhibition by means of stimulus or response generalization. To account for the sex difference in the obtained data, it was further hypothesized that although the interviews with boys' and girls' mothers did not reveal any average difference in severity of punishment, the girls, by identification, *felt* more punished than did the presumably less highly identified boys. Support for the initial hypothesis was gained from doll play, in which girls as well as boys showed a direct relationship between frequency of aggressive acts and the severity of their mothers' punishment. The data did not permit a test of the identification hypothesis.

The original study of the present subjects (Sears et al., 1957), made when they were of kindergarten age, likewise found a positive relationship between child aggression and maternal punishment. In this instance, there was no independent measure of the children's aggression; it was rated from the same interview as the measure of severity of punishment for aggression expressed toward the parents, and referred to aggression *in the home* only. However, with these data, there was no evidence of a curvilinear relationship in the girls, and we were forced to conclude either that the previously noted curvilinear relationship was a sampling error, or that the inhibitory influence of severe punishment was evident only in the school and not in the home. Theoretical reasons for the latter interpretation are not immediately obvious.

A second maternal variable shown to be positively related to aggression in the home, in this same study, was permissiveness for expression of aggression toward the parents. Permissiveness was defined as an attitude of acceptance and expectation that a child will want to be aggressive. Although permissiveness and punishment were negatively correlated with one another, both were positively correlated with child's aggression in the home.

Aggression at age 5. These findings from the total original population are accurately represented in the follow-up sample. In Table 6 are shown the correlations, separately by sex of child, between 9 available child measures, including the five aggression scales taken at age 12 and 4 of the measures taken at age 5 (dependency, aggression, conscience, bedwetting), and 12 of the measures of maternal attitude or behavior obtained from the interviews. Attention is called to the *r*'s between the age 5 measure of aggression and the measures of permissiveness and punishment for aggression toward the parents. In the original data on 379 cases, aggression at age 5 (III, 29) correlated with permissiveness for expression of aggression toward parents (III, 30) (boys) .20 and (girls) .24, and with punishment (III, 31) (boys) .14 and (girls) .18. In the follow-up sample, the figures are similar.[8]

[8] A broader measure of the permissiveness and punitiveness aspects of maternal behavior was constructed by Maccoby and Whiting (unpublished), on the basis of Milton's (1958) factor analysis of the original interview scales, for an analysis of the antecedents of the deviation doll play performed by the children at age 5. The permissiveness measure was constructed by summing four permissiveness scales: aggression toward parents, sex, dependency, and care of house and furniture. The broader punitiveness score was composed of punishment for aggression toward parents, physical punishment, demands for child to be aggressive toward other children, and irritation with dependency. These two constructed measures are called the permissiveness factor and punishment factor scores. Correlations of these scores with the various child measures are given in Tables 6, 8, and 10, but it must be remembered that they are not independent of

TABLE 6

CORRELATIONS BETWEEN THE NINE CHILD BEHAVIOR MEASURES (five aggression scales and four age 5 ratings) AND TWELVE MEASURES OF MOTHERS' CHILD REARING PRACTICES OR ATTITUDES: FOLLOW-UP SAMPLE

Child behavior measures	Child rearing variable: Permissiveness: Aggression to parents III, 30	Permissiveness factor XVIII, 5	Permissiveness for dependency XV, 5	Punishment: Aggression to parents III, 31	Punishment factor XVIII, 6	Physical punishment III, 46	Deprivation of privileges III, 47	Tangible rewards III, 36	Use of praise III, 37	Withdrawal of love III, 53	Mother warmth III, 15	Severity of toilet training II, 51
	Girls (N = 84)											
Aggression anxiety	−.20	−.13	−.02	.24	.23	.07	−.06	.03	.06	−.17	−.04	.10
Projected aggression	−.25	−.22	.05	−.08	−.12	−.09	.04	−.03	−.01	−.03	−.01	.09
Self-aggression	−.09	−.17	−.08	.11	.10	.12	.06	.19	.03	−.40	−.06	.12
Prosocial aggression	−.23	−.12	.08	.18	.21	.22	−.08	−.04	−.02	−.14	−.13	.02
Antisocial aggression	.02	.03	.10	−.13	−.15	−.06	.02	−.02	−.09	−.10	.19	.02
Aggression: age 5	.19	−.04	−.13	.18	.34	.23	.28	.13	.18	−.07	−.33	.03
Dependency: age 5	−.03	−.08	−.14	.09	.03	−.02	−.06	.09	−.11	.33	−.16	.19
Conscience: age 5	.26	.26	.06	−.10	−.23	−.25	.11	−.04	.19	.11	−.09	−.05
Bed-wetting: age 5	−.02	−.07	−.10	.08	.17	.13	−.09	.11	−.02	−.20	−.14	−.01
	Boys (N = 76)											
Aggression anxiety	−.17	−.23	.12	.10	.04	−.03	.13	.04	−.08	.21	.09	.06
Projected aggression	.17	.06	−.12	−.16	−.10	−.12	.18	.08	−.01	.27	.06	.08
Self-aggression	−.06	−.15	−.02	.03	.22	.11	.24	.39	−.14	.48	−.07	.27
Prosocial aggression	−.09	−.20	−.10	.04	−.05	−.04	.06	−.09	−.04	.10	.09	−.03
Antisocial aggression	.18	.23	.04	−.20	−.15	−.08	.10	.04	−.14	−.01	.11	.04
Aggression: age 5	.27	.20	−.11	.11	.11	.22	.00	.04	−.02	.14	−.18	.14
Dependency: age 5	.06	−.04	−.06	.02	.16	.06	−.08	.15	.19	−.10	.06	.19
Conscience: age 5	−.13	−.20	.08	−.03	−.28	−.25	−.14	−.17	.19	.03	.15	−.01
Bed-wetting: age 5	.05	.21	−.02	−.18	−.04	−.02	−.21	−.04	.06	−.22	−.05	.12

Note.—Below each named scale is given the code number from Appendix B of *Patterns of Child Rearing* (Sears, Maccoby, & Levin, 1957). Significance levels for girls: when r = .21, p = .05; when r = .28, p = .01. Significance levels for boys: when r = .23, p = .05; when r = .29, p = .01.

These measures of antecedent-consequent relationships are low, but all are in the correct direction to indicate that amount of reported aggression in the home at age 5 was positively related to both *high permissiveness* and *high punishment*. From a theoretical standpoint, this is interpreted to mean that permissiveness represented both cue or signal instigation to aggression, and an implied reinforcement, while punishment produced instigation by frustration, and conflict induced instigation. A third role that can be played by punishment—but for which there was no evidence in the original *Patterns* study comparable to that obtained on girls in the Iowa study (Sears et al., 1953)—is that of inhibitor of the punished response. As will be seen below, this inhibitory influence is clearly detectable by age 12.

Antisocial aggression at age 12. The five aggression scales provide a more differentiated measure of the children's later aggression than did the single interview rating at age 5. Unfortunately, there is no measure directly comparable to that rating, i.e., a measure of aggression in the home. Probably the one scale most comparable to it, however, is that for antisocial aggression. The early rating attempted to measure the amount of aggression the child showed in the home, mainly that directed at the parents. Since most of the mothers disapproved of such behavior in some degree, it must be considered a major early

the separate scales of which they are composed. In general, they add little beyond what is provided by the individual scales to an interpretation of the mother-child relationships except in connection with aggression anxiety and self-aggression, and they will not be discussed elsewhere.

form of unsocialized—or antisocial—aggression. The items in the antisocial scale did not refer to the same kinds of aggressive acts, nor to the same objects of aggression, but they were meant to be representative of unsocialized and unacceptable forms of aggression at age 12. None of them included a reference to parents or the home, so the similarity is but partial. And, of course, there is the radical difference between mother's report and self-report.

One would expect that antisocial aggression would be negatively correlated with those scales most representative of more highly socialized and acceptable attitudes toward aggression. Among the present dimensions, these would be prosocial aggression and aggression anxiety. In Table 3, which shows the intercorrelations obtained from the entire tested population, this expectation can be seen to be fulfilled; for both sexes the *r*'s are negative, although that between antisocial and prosocial in girls is significant at only the .02 level. If we can assume that the total tested population gives the more reliable measure, this low relation is apparently the product of a sampling accident within the smaller follow-up sample, which contributes to the total tested population; in that sample, the relationship in girls disappears altogether, although the other three coefficients are approximately the same as for the larger population (cf. Table 4).

A second expectation might be that the rating at age 5 would be positively related to the antisocial score at age 12. This rests on the presumption of a continuity of both the child's habitual reactions and the mother's methods of child rearing. Evidently this reasoning is not sound, for in Table 4 it may be seen that for the boys the *r* is −.03 and for the girls, −.08; in other words, effectively zero for both. The implications of this discontinuity from age 5 to age 12 will be considered in a later section.

The relation of permissiveness and punitiveness to all the aggression scales can be seen in Table 6. For antisocial, punishment for aggression to parents is negatively related in both sexes, though neither of the coefficients reaches the .05 level of significance. This is the opposite direction from the relations found between these same measures and the rating of aggression at age 5. If it can be presumed that these two child measures are tapping somewhat similar instigation-action sequences, then we may suppose that, during the intervening years, the inhibitory effect of the punishment had an opportunity to be influential; no longer was the punishment for aggression simply a goad to further action—it reduced the tendency to express the punished forms of behavior.

The zero correlations between age 5 ratings and antisocial, however, show that the shift was not a simple unidirectional matter. Some children suffered inhibition—and reduced aggression—while others did not. Likewise some who had been low at age 5 remained low, some became higher at age 12. The reasons for this may involve some constitutional quality in the children, in part, or what is more likely, some unmeasured change in the mother's child rearing behavior. As will be shown below in the section on Change from Age 5 to Age 12, however, there were suggestive differences in the original rearing patterns that are consistently associated with certain of these variations. It may be said at once that the severity of punishment itself played only a limited role.

The contrast, at the two ages, in the relation of severity of punishment for aggression toward parents is presented graphically in Figure 1. The abscissa shows three degrees of punishment (III, 31) as measured from the mother interviews taken when the children were age 5. The age 5 measure of aggression was from the same interviews; the age 12 measure is the antisocial aggression scale. As would be expected from the correlations reported in Table 6, the age 5 relationship (*t* between low and high groups) in girls is significant at $<.01$, while in the boys it is nonsignificant. At age 12, the decrease for boys is significant at the .08 level, while the girls show no variations at all. The difference in trend between the two ages is significant for both sexes; the difference between the two *r*'s for each is at the .05 level.

A relation of permissiveness to antisocial aggression is suggested only in the boys, however (+.18). It may be supposed that either the mother's attitude of expecting and tolerating aggression carried over its influence from the early years, or that mothers tended to be somewhat consistent in their levels of permissiveness as the boys matured. There is no

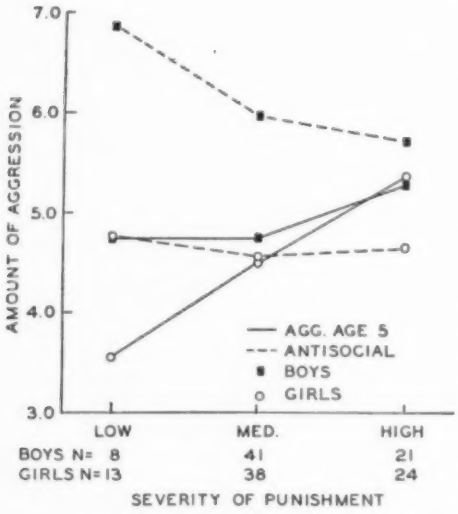

FIG. 1. Amount of aggression in the home at age 5 and antisocial aggression at age 12 in relation to severity of punishment for aggression toward parents as rated from mother interviews. (The figure is based on only those cases for which measures were available on all three variables.)

evident relationship between permissiveness and antisocial aggression in girls. The distribution of girls' scores on their mothers' permissiveness for aggression toward parents is severely skewed, however, and the r is probably an inappropriate measure of this relationship. In fact, the seven most permissively treated girls were all above the median on the antisocial aggression scale.

These various relationships may be summarized as follows: for both sexes, high punishment for aggression in the first 5 years of life was associated with a high level of reported aggression in the home, but at age 12 it was associated with a low level of antisocial aggression. In boys, at both ages, high permissiveness was associated with high aggression; in girls, this effect was clearly evident only at the earlier age, with a mere suspicion of it at the later.

In our earlier report (Sears et al., 1957, p. 401), it was shown that mothers of girls were already less permissive of aggression, when the children were but 5 years old, than were the mothers of boys. One cannot but wonder whether the stricter attitude of control toward girls may not have increased even more during the middle childhood years, and produced an essentially homogeneous level of permissiveness—or at least an even more severely truncated distribution. If this were the case, the lack of correlation between the earlier measure of the mothers' attitudes and their daughters' 12-year-old aggressive responses would be quite understandable on purely statistical grounds.

Aggression anxiety and prosocial aggression at age 12. Three possible effects of punishment have already been mentioned—instigation to aggression by frustration, and by conflict drive, and inhibition of the punished response. A fourth must also be considered—the establishment of anxiety as a response to whatever instigators are present when punishment occurs. In the present context, the anxiety with which we are concerned is elicited by the same stimuli as aggressive responses, plus the response produced cues arising from overt or imaginal or affective aggressive actions. Hence with the present instruments we would expect a positive relationship to develop between punishment for aggression to parents and aggression anxiety. Table 6 shows that the expectation is fulfilled in the girls ($p < .05$), but not in the boys. Thus far, then, one can only say that the theoretical prediction is but half-supported.

The relation of punishment to prosocial aggression is the same as to aggression anxiety. High punishment is associated with high prosocial aggression in girls, but is unrelated in boys. This positive effect of punishment is similar to that found in girls for aggression in the home at age 5; at that time, too, this relationship was nonsignificant in boys.

So, what sifts out of these rather complex comparisons is that severe early punishment for aggression tended to reduce antisocial aggression (a little more clearly so in boys) and to increase prosocial aggression (but only in girls). Nevertheless, in both sexes, prosocial aggression is as unrelated to earlier aggression as is antisocial aggression. Since both types of aggression are strongly related to aggression anxiety in both sexes—prosocial positively and antisocial negatively—it comes as a shock to discover that the earlier aggression is *positively* related to aggression anxiety in girls and *negatively* in boys. The implications of this

finding are considered below in the section on Aggression Anxiety.

Since the various attitude scale correlations referred to in this discussion of the follow-up sample are approximately replicated in the total tested population (with the single exception of the size of the negative relation between antisocial and prosocial), we are forced to the conclusion that these sex differences surrounding the early measure of aggression and the aggression anxiety measure have something to do with changes that occurred between ages 5 and 12, and with differences in the sources of aggression anxiety for the two sexes.

Mention must be made also of the role of permissiveness in connection with aggression anxiety and prosocial aggression. Both are related negatively to permissiveness, though significantly so only in the girls. Since permissiveness and punishment are rather closely related to one another (see Table 10), there is some difficulty in being certain how independent the contributions of the two measures are.

Projected and self-aggression. These two remaining aggression scales are significantly associated with each other and with aggression anxiety and prosocial aggression in boys; the same is true for girls, except that self-aggression seems unrelated to aggression anxiety and prosocial (cf. Tables 3 and 4).

The relationship to permissiveness and punishment is not entirely clear. In girls, projected aggression acts like part of the anxiety cluster with respect to permissiveness; nonpermissiveness is associated with high projection. In boys, the relation is nonsignificantly opposite. For both, punishment is nonsignificantly negative. The relations for self-aggression can best be considered later; an understanding of their antecedents will require analysis of other variables.

Conclusions. When the children's aggression at home was measured in the original mother interviews, there was evidence that both permissiveness and punishment for aggression served to facilitate such behavior. Permissiveness appeared to serve as both an instigator and a symbolic reinforcer, while punishment had its predominant effect as a further frustration and hence a goad to more aggression. The correlations that displayed these relations in the original group were significant but small; because of the reduced size of the follow-up sample, although the sizes of the correlations are similar, they are no longer statistically significant. However, the directions of the relationships in this smaller group are the same.

The more differentiated measures at age 12 appear to illustrate a rather radical shift in the role of punishment. Formerly an instigator to antisocial aggression, it now seems to have imposed an inhibitory influence on that behavior and to have created aggression anxiety. The interesting thing about the various scales is that the three nonantisocial scales—which are positively related to anxiety—seem, if anything, to enjoy a facilitating effect from punishment. It looks as though the conflict-drive effect of punishment were still operating, producing higher prosocial and self-aggression while the inhibitory influence reduced the antisocial aggression. Projective aggression, which in terms of its correlations with the other scales appears to be a part of both the anxiety cluster and the antisocial, is influenced by permissiveness in the same way as is aggression anxiety, and by punishment in the same way as is antisocial aggression.

Having only two chronological points of measurement—ages 5 and 12—one falls easily into a manner of speaking that implies the operation of one effect of punishment (facilitating) at one age and another effect (inhibiting) at the other. This is improper, of course. One can only suppose that both effects of punishment are always operative. The fact that at the earlier age the two interact in such a way as to allow the facilitative effect to be revealed in group averages, and at a later age the inhibitory effect, probably means that with the particular levels of punishment, permissiveness and other relevant variables occurring in this culture, the facilitating effects are more pronounced in early stages of socialization and the inhibitory ones do not appear so generally until later. Indeed, the concept of conflict-drive itself (Sears et al., 1953) implies the simultaneous influence of both properties.

In summary, the influence of these two child rearing variables may be hypothesized to have been as follows: high punishment for aggression reduced antisocial aggression, but

increased aggression anxiety and prosocial aggression, and high permissiveness increased antisocial aggression, but reduced the other two. Only 2 of the 12 correlations on which these statements are based reach the .05 level of significance, but the remaining 10 nonsignificant *r*'s are all in the right direction, and 5 of them are at the .10 level of significance. The evidence overall, in other words, is of just borderline significance, and though these hypotheses are congruent with the theoretical expectations concerning the influence of punishment and permissiveness, the conclusions must be accepted with considerable caution.

Alternative Responses: The Change from Age 5 to Age 12

One of the thorniest problems in aggression theory is that of alternative responses (cf. Berkowitz, 1958). It is really two problems, although these are often fused—and confused—into one. The first is catharsis. Both the instinct view and the frustration-drive concept assume a certain quantum of internal instigation that requires expression. If overt aggression is prohibited, then the impulse will be displayed in nonovert or fantasy or some other form of aggression not expected to elicit punishment (or counteraggression). This "discharge" (Freud, 1922), or "goal response" (Dollard et al., 1939), serves to reduce the strength of the residual impulse, thus, decreasing the probability of further aggressive behavior immediately thereafter. The term *catharsis* refers to this letting loose of drive instigated behavior, the lowering of tension, and by analogy to the original use of the term in connection with hysterical symptoms, the reduction of alternative expressions. An instinct theory that made no room for individual differences in the strength of the impulse would predict a negative correlation between the strengths of two temporally adjacent aggressive acts. A behavior theory (Dollard et al., 1939) that did assume individual differences—on the basis of either constitutional factors or stimulus input variation—would make more equivocal predictions, because if the *range* of individual differences were large enough, in comparison with the size of the cathartic reduction, a correlational measure could be positive. Obviously correlational methods are inappropriate for this problem. Empirical investigation of this hypothesis has so far led to quite ambiguous results, but they need not concern us here, for the present data are irrelevant to the catharsis aspect of alternative responses.

The other aspect has to do with traits, or habitual forms of aggressive action. Regardless of which of the two foregoing views concerning internal instigation is chosen, there is a clear question as to whether the inhibition of one form of aggression is likely to be associated with the stronger development of others. The time interval implied in this question is a long one—months and years—in contrast with that implied in the catharsis hypothesis. Whereas catharsis refers to an action-theory construct of immediate drive reduction and ensuing performance strength, the habitual-form question relates to the learning process and the factors determining choice of expressive form.

With respect to this issue also, there are two possible positions. They are alternative hypotheses concerning the sources of instigation to aggression. One is a drive theory, including both the instinct and the frustration conceptions of the drive's origin. The other is a cue stimulus theory. The drive theory would suggest that observed differences among people in their choice of forms of expression (e.g., antisocial, prosocial, self, etc.) are a function of reinforcement and inhibition, but that one major source of instigation to any given act is a drive. This drive could be constitutional or it could be produced by frustration of any act, including aggression. Given a frustration, then, the drive would be increased in strength and would, thus, strengthen whatever responses did occur. Concretely, it would be predicted that if one form of aggression (e.g., antisocial) were repeatedly punished, and thereby inhibited whenever external or internal stimuli aroused it, alternative forms of aggressive behavior (e.g., prosocial) that did *not* meet with punishment would be stronger than they would be if the first form (antisocial) were not inhibited.

The cue stimulus theory, on the other hand, is based on the assumption that particular forms of aggression are specifically reinforced

or inhibited, and that the strength of any one form is causally related to the strength of others only by stimulus and response generalization. Punishing one form would tend to depress other forms, on the average, but positive or negative correlations among different forms would be assumed to result primarily from similar or different schedules of reinforcement the occurrence of which would be a function of consistencies or inconsistencies in the actions of reinforcing agents (e.g., mothers). To be concrete again, we would expect, from this position, that because there in a demonstrable consistency of maternal handling of different forms of aggression, the strengths of alternative forms of aggression would tend to be positively correlated regardless of the prolonged inhibition of any one of them.

The data from the present study provide an opportunity to make one useful test of these two hypotheses. All children express *some* aggression toward their parents, and may be presumed to develop a measurable degree of habit strength for such changeworthy behavior (Sears et al., 1957). They also learn to expect gratification because aggression provides a certain control over the parents. Nevertheless, infantile forms of aggression must be relinquished eventually, and hence we may assume that by age 12 all the children, in varying degrees, had suffered inhibition of their changeworthy forms of aggression. That is, since nearly all the mothers had expressed some dislike of aggression when the children were 5, we can assume a continuing pressure against its expression, even though we have no measure of the mothers' behavior during these intervening years.

The drive theory predicts that children who were still highly aggressive at age 5 but who suffered inhibition of this behavior, and relinquished it, should show higher degrees of alternative forms than children who did not relinquish it. The cue stimulus theory predicts the opposite; i.e., the inhibited group should show lower degrees of the alternative forms.

These differential predictions can be tested by comparison of the mean scores on projected, self, and prosocial aggression obtained by two subgroups of children who did and did not, respectively, show the presumed inhibitory effects of punishment. These are children who were all High (H) (upper half of the distribution) on aggression at age 5, and then were found to be either H or Low (L)—same defining basis—on antisocial aggression at age 12.

From a developmental standpoint all these children may be considered as having posed a problem of control to their parents. They had been more severely punished, on average, and also treated more permissively than the other half of the group of children. We do not know how the mothers coped with this problem in the intervening years, but we do know that, at age 12, 29 boys and 23 girls were in the lower half of the distribution on antisocial aggression, while 17 of each sex were in the upper half. One might say that the mothers had coped successfully with the former and unsuccessfully with the latter.

Children at age 5. There were no differences at all between the means of these two groups' reported dependency, aggression in the home, or conscience. There was one difference in the boys, however, that is relevant to aggression. On the mother rating scale (III, 27) called "extent to which parent has encouraged child to fight back," there was a nonscalar code labeled "Problem hasn't come up; child has always defended self." Of the 17 High-High (HH) boys, 7 were so judged; of the 29 High-Low (HL) boys, only 2 were. The difference by χ^2 is at $p < .01$. Only 1 of the 7 HH boys was treated as nonpermissively as the 2 HL boys with respect to expressing aggression toward other children. The boys who did not reduce their aggression over these years were permitted to—and did—defend themselves against other children by fighting, even at age 5.

Children at age 12. As is shown in Figure 2, the differences in test measures for the boys are much greater than for the girls, although both are in the same direction. They tend to support the drive theory, though not unequivocally. The boys whose aggression had decreased (Group HL) were higher, by t test, on aggression anxiety ($p = .001$), self-aggression ($p = .10$), and prosocial aggression ($p < .001$). For girls, the equivalent values are $p < .05$, *ns*, and .15. There was no difference on projective aggression in either

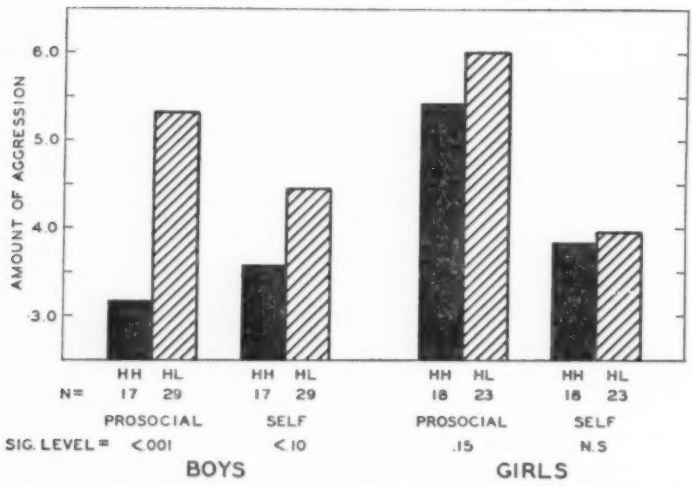

FIG. 2. Mean frequency of prosocial and self-aggression in two groups equally high on aggression at age 5 but one of which (HH) was high on antisocial aggression at age 12 and the other of which was low (HL).

sex. These differences, particularly those with respect to anxiety and prosocial, are what might be expected from more severe control by both punishment and nonpermissiveness, although the anxiety measure is not one of the alternative forms we are considering here, of course.

Mothers' behavior when children were age 5. The evidence for any such difference of treatment in the early years is minimal, however. For both sexes, there was slightly less permissiveness for aggression toward parents and toward other children, but none of these four differences reached the .05 level of significance. There was no difference in severity of punishment for aggression toward parents, and only a slight nonsignificant difference on physical punishment, with the HL group having suffered the more severely. The latter group had also been less strongly encouraged to fight back when attacked by other children.

These nonsignificant differences are all in the right direction to support the inference that the group of originally aggressive children who *did* drop into the lower half of the distribution on antisocial aggression were ones whose mothers were initially more controlling and less inciting of aggression. But the differences in maternal behavior are nowhere near as great as the collateral differences in aggression anxiety and prosocial aggression might lead one to expect.

It is worth adding, perhaps, that none of the following maternal variables showed any differences whatever between these two groups, in either sex: severity of toilet training, warmth, use of the four disciplinary techniques of tangible rewards, praise, deprivation of privileges, or withdrawal of love, not following through on threats of punishment, or who did the disciplining in the family.

These findings suggest that the effective suppression of an aggressive child's early expression of aggression in the home tends to increase the strength of prosocial and self-aggression in the later years. This conclusion is in accord with the drive theory of instigation and in opposition to the cue stimulus theory. What little information can be gained from the mother interview scales suggests, further, that nonpermissiveness rather than punishment was the technique used for controlling the aggressive behavior.

Aggression Anxiety

In the discussion so far there has been repeated reference to sex differences in the sizes of relationships between the permissiveness-punitiveness antecedents and certain of the aggression scale consequent measures. These differences have been glossed over and treated as if they were possibly no more than variations in sampling. There are certain differences in connection with aggression anxiety that

TABLE 7
AGGRESSION ANXIETY AND SELF-AGGRESSION: CORRELATIONS WITH CHILD BEHAVIOR VARIABLES MEASURED AT AGE FIVE[a]
(Girls: $N = 84$; Boys: $N = 76$)

Child behavior: age 5	Code number	Aggression anxiety		Self-aggression	
		Girls	Boys	Girls	Boys
Aggression	III, 29	.18	−.21	.04	−.10
Dependency	III, 10	.04	−.16	.05	.03
Conscience	III, 34	−.24	.30	−.07	.00
Bed-wetting	II, 50	.00	−.17	.24	−.18
Quarreling with sibling	III, 21	.04	−.19	.00	−.08

[a] See also Table 4.

can no longer be ignored, however. The most curious ones are illustrated in Table 7, which shows a sharply defined sex difference in direction of correlation between aggression anxiety at age 12 and both aggression and conscience at age 5. Both these correlational differences are significant at $p < .01$.

What is surprising about this finding is that there is only one significant sex difference among the correlations *within* either of the two sets of child measures made at the respective ages. As can be seen in Table 4, the six correlations between (age 5) aggression, dependency, conscience, and bed-wetting are in the same direction and of about the same size for the two sexes except for that between conscience and bed-wetting, and the difference of .12 between the two nonsignificant correlations of aggression with dependency. Likewise, among the five aggression measures taken at age 12, only 1 of the 10 possible comparisons is significant at the .05 level, namely, that between the two correlations of antisocial and prosocial aggression mentioned earlier; for the boys, the r is −.36, and for the girls, −.03. None of the r's relating aggression anxiety to the other four aggression measures is significantly different for the two sexes. Furthermore, aggression anxiety is almost identically related (+.34 and +.37) with the measure of femininity obtained, at age 12, by the self-administered M-F test derived from Gough's CPI Fe scale. Paralleling this intrasex evidence of the "feminine" quality of aggression anxiety is a highly significant sex difference in the means on this scale (see Table 5).

Now as to antecedents. The predictable effect of punishment as a determiner of aggression anxiety in girls has already been mentioned. It is to be noted, however, that only punishment for aggression toward parents, and the punishment factor score, are so related (see Table 8). Physical punishment and punishment for dependency are not. Neither are the various techniques of discipline: praise, use of tangible rewards, deprivation of privileges, or withdrawal of love. Indeed, the last is somewhat negatively related (−.17).

None of the punishment measures shows a relation to aggression anxiety in boys except withdrawal of love, which surprisingly correlates +.21.

Nonpermissiveness, however, is associated with aggression anxiety in both sexes (Table 8), but *only* with respect to aggression toward

TABLE 8
AGGRESSION ANXIETY AND SELF-AGGRESSION: CORRELATIONS WITH CHILD REARING VARIABLES[a]
(Girls: $N = 84$; boys: $N = 76$)

Child rearing variable	Code number	Aggression anxiety		Self-aggression	
		Girls	Boys	Girls	Boys
Permissiveness:					
Aggression to parents	III, 30	−.20	−.17	−.09	−.06
Aggression to other children	III, 28	.27	.14	−.07	−.13
Aggression to siblings	III, 22	.04	−.08	.11	−.19
Dependency to mother	XV, 5	−.02	.12	−.08	−.02
Factor score	XVIII, 5	−.13	−.23	−.17	−.15
Punishment:					
Aggression to parents	III, 31	.24	.10	.11	.03
Dependency to mother	XV, 7	.13	−.13	.10	.23
Physical punishment	III, 46	.07	−.03	.12	.11
Factor score	XVIII, 6	.23	.04	.10	.22
Child encouraged to fight back	III, 27	.23	.07	.01	−.15
Reward for dependency	XV, 6	−.18	.03	−.20	−.15
Withdrawal of love	III, 53	−.17	.21	−.40	.48
Use of praise	III, 37	.06	−.08	.03	−.14
Use of tangible rewards	III, 36	.03	.04	.19	.39
Deprivation of privileges	III, 47	−.06	.13	.06	.24
Severity of toilet training	II, 51	.10	.06	.12	.27
Sex role differentiation	III, 20	−.08	−.06	.15	.10
Mother warmth	III, 15	−.04	.09	−.06	−.07
Father warmth	III, 65	.09	−.15	−.14	−.17
Amount of father caretaking: infancy	II, 18	.11	−.12	.02	.02
Amount of father caretaking: age 5	III, 64	.28	−.04	−.12	−.08
Which parent disciplines (father = 1)	III, 66	−.23	−.06	−.10	−.16

[a] See also Table 6.

parents (or as measured by the factor score). Exactly the opposite relation obtains with permissiveness for aggression toward other children (excluding siblings); aggression anxiety is slightly but nonsignificantly related to this scale in a *positive* direction for boys and significantly at $p < .01$ level in girls. This reversal in the role of permissiveness is the key to a possible explanation of the sex difference.

A glance down the boys' column of Table 8 shows but two significant *r*'s (.05 level or better), namely, those with the permissiveness factor score and with withdrawal of love. In the girls' column, however, there are six that reach that level, and three others (including withdrawal of love) close to it. These mother scales are by no means independent of one another, of course, and the actual numbers of significant *r*'s are less important than the kinds of variables that seem to be most connected with aggression anxiety.

The hypothesis is suggested that aggression anxiety has quite different dynamic and genetic qualities for boys and girls. In boys, it is a part of the conformity syndrome, a close relative of conscience, and one of the consequences of relatively high but nonpunitive pressure toward socialization; it is a positive reaction to the same love oriented method of discipline and control that succeeds at an early age in reducing both dependency and aggression toward parents and quarreling among siblings. It is a feminine quality only in the sense that most behavioral representations of rapid and effective socialization in the preschool years are more characteristic of girls and women than of boys and men. It is established by a general aura, in the home, of nonpermissiveness for infantile and destructive impulse expression; the sanctioning power when deviant behavior does occur comes from the threatened loss of the love relationship to the mother. Dynamically, it is one of many responses to stimulation that arouses affiliation desires and conformity behavior.

In girls, aggression anxiety is more complex and has a quite different genesis. Dynamically it is part of the syndrome of forced conformity and subordinated autonomy in which anxiety is evoked by any instigator, internal or external, that serves to arouse incipient aggressive actions or feelings. It is the product of an internalized Type II conflict (Sears & Hovland, 1941) in which the positive response is a forbidden form of aggression, the negative is fear of punishment, and the source of instigation is a frustration induced drive.

This conception is in sharp contrast to that about boys, suggested above. The reasoning that favors some such difference rests on the quantitative reversal of the relationship between aggression and aggression anxiety in the two sexes. Constitutionally, girls are probably less overtly aggressive than boys, and certainly the conditions for their rearing are such as to produce more rapid maturing and relinquishing of earlier changeworthy behavior, aggressive and otherwise. They evidently receive fewer external instigators to aggression and have fewer response produced cues. Girls as a generality, then, should show less aggression anxiety than boys if both sexes have such responses evoked mainly by internal instigations to aggression. On the contrary, girls show much more.

Genetically, the conditions that establish aggression anxiety in girls may be summarized under the label of conditions *antithetical to appropriate sex typing*. With the exception of nonpermissiveness for aggression to parents (which *is* appropriate), all the larger correlations in the girls' column of Table 8 are with items of child rearing that either were shown originally (Sears et al., 1957, p. 401) to be different for boys and girls or are related to father involvement and to the punishment of aggression or dependency. The direction of every one of these *r*'s is opposite to that implied by the sex difference in rearing. Thus, for example, girls are less frequently encouraged to fight back when attacked, but the correlation between this scale and aggression anxiety is positive; likewise, girls are more commonly disciplined by withdrawal of love, but this scale is negatively related to aggression anxiety. The amount of father caretaking at age 5 was not different for the two sexes in the original population, nor is it significantly so in the follow-up sample, but proportion of father discipline was; it seems reasonable to suppose that greater maternal care for girls and paternal participation with boys is appropriate for sex typing in each case. Both these scales show a positive rela-

tionship between father participation and aggression anxiety in girls.

Not all of the sex differences in rearing experiences described in the original study are represented in the list of larger correlations with aggression anxiety. Time for completion of weaning, warmth to infant, use of maids and sitters in infancy, kinds of chores required, expectancy of going to college, amount of use of praise and amount of physical punishment, for example, do not show the relation. Neither does sex role differentiation by the mother, the scale measuring the degree to which the mother conceived that there were differences between the sexes that required differences in rearing. One might expect that those mothers who were not aware of sex role differences might treat their daughters less appropriately for feminine sex typing and, thus, produce more aggression anxiety. The r of $-.08$ is of insufficient size to be said to support this expectation, however.

To summarize, this conception of the different dynamics and genetics of aggression anxiety in the two sexes is congruent with the age 5 correlations shown in Table 7. There it is seen that, for girls, high aggression anxiety is associated with higher aggression in the home and lower conscience. There is no relation to dependency or quarreling among siblings, as there is among the boys. For the boys, on the other hand, the anxiety is associated with low aggression and high conscience. On the antecedent side (Table 8), the different dynamics suggested for the sexes are congruent with the more specifically aggression connected nonpermissiveness in the girls' experience and the more general (factor) nonpermissiveness in the boys', and with the clear influence of punishment in the girls' and its irrelevance in the boys' experience. The other relevant child rearing dimensions that correlate significantly with the anxiety measures in girls are ones which have been shown previously to be relevant to sex-typing, and the direction of correlation in every instance is such as to associate high aggression anxiety with the inappropriate sex-typing practice. The theory also accounts for the greater aggression anxiety of girls than of boys, even though girls are less aggressive in both overt behavior and fantasy (P. S. Sears, 1951); the explanation of this phenomenon is not in terms of a greater frequency of inappropriate sex typing experiences in girls, of course, but is seen as a consequence of the development of internalized Type II conflict.

TABLE 9
AGGRESSION ANXIETY AND ADULT ROLE: CORRELATIONS IN TOTAL TESTED POPULATION AND FOLLOW-UP SAMPLE

Adult role measures	Total tested population		Follow-up sample	
	Girls	Boys	Girls	Boys
Enforcing rules	.17	.18	.15	.25
Accepting rule enforcement	.19	.24	.10	.36
Giving nurturance	.02	.20	−.04	.37
Adult versus child choice	.22	.28	.23	.36
Number of cases	264	261	84	76

One further set of differential relationships should also be demonstrable. These have to do with maturity of role choice and nurturance. One would expect that boys would show substantial positive correlations of aggression anxiety with the mature and nurturant forms of behavior, while the girls would show no relationship. Table 9 gives evidence for these propositions in the form of correlations between aggression anxiety and four measures reported elsewhere by Maccoby (1961). These measures relate to the positive choice of adult-like roles in connection with *using rule enforcement*, *accepting rule enforcement*, *giving nurturance*, and a general series of items having *subordinate-superordinate role choice comparisons*. In the total tested population the girls do have significant r's with respect to all but the giving nurturance scale, though the boys' r's are slightly larger. In the follow-up sample, the sex differences are still larger. On giving nurturance, both sets of figures show a significant positive relation ($p < .01$) in boys and zero relation in girls, as expected. These figures support the expectation so far as nurturance is concerned, but do not with respect to the other forms of mature behavior.

Conclusions. It would be an obvious absurdity to suppose that either of these dynamic or genetic conceptions of aggression anxiety is applicable rigidly to one sex only. Undoubtedly many males have had their aggression anxiety somewhat exacerbated by parental punishment for aggression or by too severe demands for aggressive behavior in

their relations with other children. The virtually zero correlations between these dimensions and boys' aggression anxiety cannot be interpreted as meaning that those variables have had no influence for *any* boy in the group, but simply that as a cumulative matter within the whole matrix of influences, these are not detectable for the group as a whole. Likewise, in the female, one must assume that the affiliative conforming motivation created in part by nonpermissiveness and by appeal to maturity and conscientiousness is also a determinant of aggression anxiety. The relations with the adult role scales favor this interpretation. Indeed, one cannot but wonder whether both sexes may not about equally derive aggression anxiety in this way, but that girls have overlaid a more severe degree of it by those socializing experiences which, in the exciting of aggressive impulses counter to girls' proper nonaggressive sex role, are inappropriate to the sex-typing process.

In any case, the evidence presented here suggests two quite different dynamic and genetic forms of aggression anxiety. Whatever combinations of them may occur in individual cases, within the limits of our sample, one is predominant in boys and the other in girls at age 12.

Self-Aggression

Although the internal consistency of the self-aggression scale is virtually nil, the sizes and directions of its relationships with the other child measures are so consistent, both between the follow-up sample and the 1958 classmates and, with one exception, between boys and girls, that its child rearing antecedents seem worth investigation. The one exception is in respect to aggression anxiety; for boys, these two scales are positively related in both groups (follow-up, $r = .25$, and total tested population, $r = .24$), while for girls the slight positive relation is nonsignificant in both groups (respectively, .11 and .08). In view of the different dynamics of aggression anxiety described in the preceding section, this is scarcely surprising, but does warrant an attempt at interpretation. The other relationships of self-aggression are also a little smaller for girls than for boys, but not significantly so. The r with projected aggression is clearly positive in both sexes, while the r's with both prosocial and antisocial aggression are effectively zero.

From a theoretical standpoint (Dollard et al., 1939), self-aggression has been considered a displacement of an aggressive response from the frustrating agent to the self, produced by anticipated punishment for direct aggression toward the agent. Such an expectancy would arise if there had been repeated punishment for being aggressive toward the frustrator. Since aggression toward the self is intrinsically painful, this would be the direction of aggression adopted only if most other objects of displaced aggression were also interdicted.

In the present context, we would expect that high self-aggression would be associated with various child rearing experiences that both heightened frustration and served to prevent the expression of overt aggression toward parents, siblings, and other children. In Table 8 are shown the relevant correlations for a test of this hypothesis. It is apparent at once that the boys' self-aggression is significantly related to more socialization variables than the girls'. Indeed, the only significant correlation for the latter is with the nonuse of withdrawal of love as a disciplinary technique.

For the boys, however, there is substantial support for the hypothesis. First, we would expect any early socialization experiences that were strongly frustrating to create high instigation to aggression and, at age 5, high aggression toward the frustrator. In a previous study (Sears et al., 1953), there was some evidence for a positive connection between severe toilet training and aggression in the preschool; in the present follow-up sample of boys there had been a nonsignificant relation ($r = .14$) between severity of toilet training and aggression in the home. Severity of toilet training was correlated .41 with use of physical punishment, which suggests that the hypothetical conditions outlined above were fulfilled—namely, high frustration, with some consequent aggression, and punishment for expressing aggression. As was predicted, there is a significant relation ($r = .27$) between severity of toilet training and self-aggression. None of the other aggression scales is significantly related to this antecedent.

A second source of early frustration was the existence of siblings; we may assume that

TABLE 10

INTERCORRELATIONS AMONG TWELVE CHILD REARING MEASURES OBTAINED FROM THE MOTHER INTERVIEWS: FOLLOW-UP SAMPLE

(Girls' mothers: $N = 84$; Boys' mothers: $N = 76$)

Girls / Boys — Child rearing variable	Code number	Item	1	2	3	4	5	6	7	8	9	10	11	12
Permissiveness: Aggression to parents	III, 30	1		.70	.24	−.44	−.28	−.30	−.04	−.01	.19	.04	.15	−.29
Permissiveness factor	XVIII, 5	2	.67		.64	−.40	−.35	−.35	−.13	−.07	.16	−.07	.27	−.36
Permissiveness for dependency	XV, 5	3	−.05	.43		−.22	−.35	−.09	−.12	.00	.05	−.14	.35	−.10
Punishment: Aggression to parents	III, 31	4	−.61	−.47	.08		.62	.58	.05	.15	.00	−.02	−.37	.18
Punishment factor	XVIII, 6	5	−.31	−.22	−.04	.61		.70	.10	.18	.00	.07	−.45	.18
Physical punishment	III, 46	6	−.22	−.12	−.04	.34	.68		.08	.22	−.05	−.09	−.34	.21
Deprivation of privileges	III, 47	7	−.07	−.04	.10	.06	.14	.25		.33	.22	−.02	−.11	.20
Tangible rewards	III, 36	8	.01	−.17	−.07	.13	.16	.16	.22		.02	−.08	.07	.21
Use of praise	III, 37	9	.10	−.07	−.13	−.04	−.10	−.12	−.09	−.05		.12	.14	.21
Withdrawal of love	III, 53	10	−.19	−.27	−.16	.36	.34	.41	.34	.32	−.20		−.20	.08
Mother warmth	III, 15	11	.08	.23	.40	−.01	−.24	−.30	−.08	−.06	.15	−.49		−.18
Severity of toilet training	II, 51	12	−.23	−.46	−.26	.31	.34	.41	.03	.21	−.05	.30	−.37	

Note.—Girls' mothers above the diagonal; boys' mothers below.

other children in a family, either older or younger, add not only companionship but also competition for things, space, and people. Here again there is a detectable relationship; by *t* test, *only* boys were less self-aggressive than those with siblings ($p = .04$).

A third source of frustration at the earlier age was punishment for aggression and dependency. However, punishment should also have inhibited aggression eventually, so that on two counts a high level of punitiveness would be expected to be positively related to self-aggression. Interestingly enough, punishment for aggression toward parents shows no relation at all ($r = .03$) and physical punishment a nonsignificant amount (.11), while punishment for dependency (.23) and the punishment factor score (.22) are both significantly related. Furthermore, two of the techniques of discipline—which are indirect measures of punishment—are associated in the same way, deprivation of privileges (.24) and withdrawal of love (.48). The use of tangible rewards, which is positively related to deprivation of privileges, is also related positively to self-aggression (.39). (In boys, the four disciplinary techniques have a median intercorrelation of .28, while in girls, the median is .05; see Table 10.)

As is the case with nearly all the other antecedent-consequent relations that have been reported in this and similar studies, these correlations are small on an absolute scale. This is to be expected, of course, in view of the multiple antecedents responsible for a given item of behavior; no one variable contributes much of the total variance. The influence of these several types of experience can be displayed more dramatically, however, if comparison is made of a few extreme cases. For example, if the upper and lower halves of the distributions on permissiveness and punishment factor scores are combined as follows:

High permissiveness, low punishment ($N = 17$)

Low permissiveness, high punishment ($N = 23$)

the mean scores on self-aggression for the two subgroups are different at the level, $p = .02$,

one-tailed test,[9] with the second subgroup the higher.

Of course, with the relatively low intercorrelations among the child rearing variables (Table 10), adding further variables to such a crossbreak rapidly reduces the number of cases. For example, adding severity of toilet training as a third selector to the above crossbreak gives the following:

High permissiveness, low punishment, low severity of toilet training ($N = 10$)

Low permissiveness, high punishment, high severity of toilet training ($N = 17$)

The second group has higher mean self-aggression at the level, $p < .01$, one-tailed test. If one uses four variables instead of three for obtaining some cumulative effects on self-aggression, the number of cases becomes very low indeed. However, the second of the following pair of subgroups has higher mean self-aggression at the level, $p < .004$, one-tailed test, even with so few cases; the range of scores for the first group is from 1 to 4 and that for the second group is 4 to 9:

High permissiveness, low punishment, low deprivation of privileges, low use of tangible rewards ($N = 6$)

Low permissiveness, high punishment, high deprivation of privileges, high use of tangible rewards ($N = 8$)

By the time a fifth variable is added, the N becomes so small as almost to nullify the advantages of dealing with extreme cases. For instance, if severity of toilet training is now added to the above crossbreak, the Ns for the first and second groups are only 2 and 6, respectively, and the significance of the difference between means is $p = .02$, one-tailed test.

In the foregoing comparisons, permissiveness has been added to the several punishment measures. As Table 8 shows, neither permissiveness for dependency nor for aggression toward parents is at all related to self-aggression, but there are nonsignificant negative correlations with permissiveness for aggression toward siblings (−.19), toward other children (−.15), the permissiveness factor (−.15), and encouragement to fight back when attacked (−.15). In other words, nonpermissiveness in every direction (except towards parents) seems related to self-aggression. If we use the same device for selecting extreme cases of nonpermissiveness alone, as was illustrated above in connection with punishment plus nonpermissiveness, the following tiny subgroups are shown to be significantly different at the level, $p = .05$, one-tailed test; the second group's being the higher, of course:

High permissiveness, low punishment, high permissiveness for aggression to other children, high encouragement to fight back ($N = 3$)

Low permissiveness and high punishment on same scales ($N = 7$)

Restriction on early expression of aggression. According to the theory (Dollard et al., 1939), self-aggression should develop most vigorously when the early expressions of aggression are successfully inhibited. There are three ways of testing this proposition in the present data, and all the appropriate comparisons show correlations or mean differences in the predicted directions, though the levels of significance are not very convincing.

1. The comparison of age 5 aggression with antisocial aggression offers one test. The 29 boys who were judged high at age 5 but low at age 12 expressed more self-aggression than the 17 boys who remained high at age 12 ($p = .05$, one-tailed test). It will be recalled that the permissiveness factor score was significantly lower for the more inhibited of these subgroups; so this present rather tenuous difference can be reasonably interpreted as related to the restriction produced by such nonpermissiveness.

2. A second test involves toilet training and bed-wetting. In the earlier study of these children, it was shown that severe toilet training was positively related to ineffective control of urination. In the present follow-up sample this correlation is only .12, however, and there are an almost exactly equal number of boys who had severe toilet training and who did and did not suffer from enuresis at age 5. If we may interpret the enuresis as a form of disguised aggression, then the 13 boys who showed such behavior may be considered less restricted in early aggression than the 12 boys who did not suffer enuresis. As predicted, the former had lower self-aggression

[9] Significance values by one-tailed test refer to hypotheses formulated before the data were collected, and which were derived directly from the more general theory published 19 years earlier (Dollard, Doob, Miller, Mowrer, & Sears, 1939, Ch. 3).

scores than the latter, but the difference does not quite reach the .10 level, one-tailed test. Now, if we add the differential effects of high and low frustration to this comparison, as in the following pair of groups, we find a much larger difference in the predicted direction:

High bed-wetting, low severity of toilet training ($N = 12$)

Low bed-wetting, high severity of toilet training ($N = 12$)

The second group—with restricted early aggression and high level of frustration—was more self-aggressive at age 12, $p < .001$, one-tailed test, than the first group, which had *un*restricted aggression (in the form of bed-wetting at age 5) and a relatively low level of frustration (nonsevere toilet training).

3. A third test takes us back to the most highly controlled subgroup of boys as measured by the factor scores for permissiveness and punishment. These boys were subjected to low permissiveness and high punishment. The restriction hypothesis predicts that those who were already low in their expression of aggression at age 5 ($N = 6$) would show greater self-aggression at age 12 than those who were still high at age 5 ($N = 5$); for this comparison, the 10 boys rated on the middle point of the nine-point aggression scale at age 5 are disregarded. As predicted, the restricted group does show higher self-aggression than the nonrestricted group, but $p = .20$. If we add to this comparison the congruently extreme groups on amount of quarreling with siblings that the mother reported, we get too few cases in the cells to permit further testing. However, it is worth noting that the boy who obtained the highest self-aggression score in the entire follow-up sample was not only in the least aggressive-in-the-home-at-age-5 group but also received the lowest possible rating on amount of quarreling with siblings, and was in the earliest toilet trained group although he suffered the third highest severity of toilet training. In other words, he was not only in the least permissive and most punitive home atmosphere, but he suffered very severe frustration very early, responded to it by submission, and was highly restricted (inhibited) in his aggression at age 5.

These various bits of evidence support the proposition that, in boys, restricted expression of aggression in the earliest years is associated with a higher level of self-aggression in middle childhood, but the statistical significance of the differences cited is too low to give great confidence in their replicability. Nevertheless, these findings, coupled with the predicted influences of punishment and nonpermissiveness, do give one an optimistic sense of rightness about the general theory that self-aggression is a displacement of strong aggression from frustrating parental agents to the self as a consequence of anticipation of punishment for expressing aggression directly toward the frustrators or other immediately available objects.

Aggression anxiety. It will be recalled that the correlation between aggression anxiety and self-aggression was the only one among the age 12 scales that was significantly different for the sexes; the boys' r was higher than the girls'. The findings reported above suggest a reason. In boys, nonpermissiveness is prominent among the child rearing antecedents for both responses; hence a positive relation between the two is to be expected. But for girls, what little suggestion there is of meaningful antecedents to self-aggression lies nonsignificantly in this same area (e.g., low permissiveness factor) and in only one narrow aspect of the antifeminine sex-typing area (low reward of dependency and low use of withdrawal of love). The other relevant antecedents to girls' aggression anxiety are unrelated to self-aggression. Thus, it is to be expected that these two scales would themselves be but weakly correlated in girls.

Role of the father. So far we have been concerned mainly with theoretical properties of behavior, such as punishment, permissiveness, and restriction (inhibition) of action. However, the measures of these dimensions, with the exception of the factor scores, have referred to particular substantive dimensions, e.g., permissiveness for aggression toward parents, or toward other children, punishment for dependency, physical punishment, etc. Furthermore, all of these measures imply, even when they do not explicitly state, an agent of the action as well as a substantive quality, e.g., punishment (*by mother*) for dependency (*toward mother*). There are, in other words, three aspects to every child rearing dimension: a property that represents a theoretical variable in the general theory of

action and learning (e.g., punishment); a substantive property that describes a dimension within the special theory of human social interaction (e.g., dependency, aggression); an agent, also defined within this latter theory, who is the opposite member of a dyad in which the child is object (e.g., mother, father).

In the discussion of permissiveness and punitiveness as determiners of antisocial aggression, aggression anxiety, and self-aggression, we have been concerned with the first of these three aspects. Evidence concerning the effects of these two antecedents was sought wherever it might be found, regardless of the substantive dimension being measured; sometimes the factor score gave the strongest correlation, sometimes one of the more specific dimensions. When the antecedents of aggression anxiety came under consideration, however, the marked sex differences required more careful consideration of the second and third aspects, the substantive quality of the dimensions involved and the agents; conclusions concerning the sex-different dynamic and genetic characteristics of aggression anxiety were stated in terms of the sex appropriateness of socialization procedures. This kind of principle belongs in the special theory of human social interaction, a theory that in its present degree of development is certainly "special" to human behavior and may be more narrowly limited than that; recent efforts to define transcultural variables are promising, but their success has not yet been fully assessed (Sears, 1961; Whiting & Child, 1953).

Any consideration of the role of specific persons in socialization belongs to the special theory. In connection with girls' aggression anxiety, evidence was presented that high *father* participation in a girl's rearing at age 5 was conducive to high anxiety. There is some indication that fathers also have a special influence on the development of self-aggression in boys. The evidence will be described but briefly, and only for purposes of suggesting an hypothesis in the realm of the special theory (where fruitful hypotheses are harder to come by than in the general theory of action and learning); for while trends within the data are quite consistent, almost none of the crucial antecedent-consequent comparisons reaches a satisfactory level of statistical significance.

It will be noted in Table 8 that two scales involving the father show small correlations with self-aggression, his coldness, and his degree of participation in discipline when both parents are present. These two measures are uncorrelated with one another, and hence it is theoretically possible to maximize their combined effects by a crossbreak as follows:

FATHER	*N*	% HIGH ON SELF-AGGRESSION
Cold		
Father chief disciplinarian	13	46
Parents about equal	12	8
Mother chief disciplinarian	17	35
Warm		
Father chief disciplinarian	14	28
Parents about equal	12	25
Mother chief disciplinarian	7	14

Regardless of whether the father is described as warm or cold, if he is the chief disciplinarian of the 5-year-old boy, the boy is more likely to have a high self-aggression score at age 12 than if his mother is the chief disciplinarian. Likewise, if the father is cold, regardless of which parent does the disciplining, the self-aggression scores are higher. None of the comparisons approaches significance by either t tests between means or by χ^2. The top and bottom groups of the list maximize these joint effects.

A third variable that may be influential is the father's strictness. In general, the parent who is stricter tends to be the chief disciplinarian (by χ^2, $p < .01$), and there is a close positive relationship (by χ^2, $p < .001$) between the father's rated strictness and his adoption of the chief disciplinary role. In order to see the effects of his strictness per se, then, the first and fourth groups above must be used for the next crossbreak.

FATHER	*N*	% HIGH ON SELF-AGGRESSION
Cold, chief disciplinarian		
High strictness	7	28
Low strictness	6	67
Warm, chief disciplinarian		
High strictness	4	75
Low strictness	10	10

In spite of the large percentage difference in the top pair of groups, the actual means are almost identical; by both t test and χ^2 the

difference is not significant. Between the strict and lenient *warm* fathers, however, the percentage difference in high self-aggression scores faithfully represents a large mean difference also; by χ^2, $p < .02$, and by t test, $p < .01$.

In summary, there are minimal indications that high participation of the father as disciplinarian, and his coldness, are both conducive to self-aggression, but only in a warm father does his strictness have a facilitating effect. This latter finding is reminiscent of our earlier discovery (Sears et al., 1957) that a mother's withdrawal of love was influential on the development of conscience only if the mother was warm to start with; she could not affect a child if she tried to withdraw something she had never given! With the fathers, one may suspect that somewhat the same principle holds—his being strict can frustrate his son only if there would be love gained in the absence of strictness.

These hypotheses are presented blind, in one sense. We do not know to what extent the effects of the father's behavior are simply reflections of the operation of the general principles of learning and action involving punishment and nonpermissiveness. It may be that fathers used more of the tactics that produced self-aggression than did mothers, and that the seeming effect of father behavior as such is but a reflection of the more intense use of them. Even if this is so, however, a principle having to do with the influence of fathers, as such, would be a useful addition to the special substantive theory of socialization (in our culture) if it could be substantiated.

Prosocial Aggression

There is little to be added, concerning prosocial aggression, to what was mentioned in the initial discussion of punishment and permissiveness. Like aggression anxiety, with which it is strongly correlated, it is significantly higher in girls than in boys (Table 5). Within the boys' group, it correlates significantly with the femininity measure, and within the total tested group of girls, likewise; within the follow-up sample of girls, however, this positive relation is reduced almost to zero. Again we are presumably dealing with sampling error.

So far as antecedents are concerned, the same high permissiveness and high punishment associated with aggression anxiety in girls are found similarly associated with prosocial aggression, though none of the other anti-sex-typing variables appears to be related. For boys, there is some slight evidence of the same nonpermissiveness that was found in aggression anxiety, and a complete lack of any other child rearing correlate.

SUMMARY

A follow-up sample of 76 boys and 84 girls from the original population of 202 boys and 177 girls described in *Patterns of Child Rearing* (Sears et al., 1957) was given a set of self-administering aggression scales at age 12. These included measures of antisocial, prosocial, projected, and self-aggression and aggression anxiety. Reliabilities of all but the self-aggression scale were satisfactory.

Intersex comparisons showed significantly higher scores for girls on aggression anxiety and prosocial aggression, and for boys on antisocial aggression. Intrasex correlations with a measure of masculinity-femininity, derived from Gough's (1957) Fe scale, were in directions congruent with these obtained sex differences in mean scores.

Among the aggression measures, antisocial aggression was negatively correlated with both prosocial aggression and aggression anxiety, and had virtually zero relation to projected and self-aggression. These latter, however, were significantly related positively to aggression anxiety. In other words, there appeared to be an anxiety cluster including four of the scales, with r's ranging from 0 to +.50, and antisocial aggression, with r's between it and the members of the anxiety cluster ranging from 0 to −.44.

Antecedents of these age 12 scores were sought among the ratings of the mother interviews that had been obtained when the children were in kindergarten (age 5). Primary attention was given to dimensions involving permissiveness and punitiveness, such as those for control of aggression and dependency and methods of discipline.

Antisocial aggression was found to be positively related to high permissiveness and low punishment. The former of these two effects was the same as that with respect to aggression in the home at age 5 (rated from mother interviews); the latter effect was just the

opposite from that found at age 5. At the earlier period punishment incited aggression, preponderantly, while at age 12 the negative correlations are interpreted as exemplifying the inhibitory influence of punishment.

Prosocial aggression and aggression anxiety likewise were found to be associated with high permissiveness, but in these instances *high* punishment was related to the two aggression measures. This was the opposite of the relation with antisocial aggression.

Evidence was presented to support an expectation, based on a theory of frustration induced drive, that the successful inhibition of early forms of aggression in severely punished children would produce heightened amounts of aggression anxiety, prosocial, and self-aggression at a later age.

An important sex difference was found in the antecedents for aggression anxiety, a fact that led to a theoretical interpretation that this phenomenon has different dynamic and genetic characteristics for the two sexes. In boys, this anxiety appears to be a part of a general pattern of rapid socialization and high conformity in a love oriented atmosphere of tight control of aggression. In girls, such anxiety appears to be conflict induced by an internalized Type II conflict derived from socialization experiences that exacerbate their aggressive activities because the preponderant training is antithetical to appropriate feminine sex-typing. This interpretation is supported by a series of antecedent-consequent relationships, as well as by a cluster of intercorrelations between aggression anxiety and earlier forms of behavior, that are almost entirely different for boys and girls.

Self-aggression in boys was found to be strongest in those who had suffered most severe frustrations and the most severe control of outward directed aggression in their early years. Early successful restriction of aggressive behavior, as in severe toilet training with early success and in low expressions of aggression toward parents and siblings, was also conducive to high self-aggression.

REFERENCES

BERKOWITZ, L. The expression and reduction of hostility. *Psychol. Bull.*, 1958, **55,** 257–283.

CLARK, R. A. The effect of sexual motivation on phantasy. In D. C. McClelland (Ed.), *Studies in motivation*. New York: Appleton-Century-Crofts, 1955.

DOLLARD, J., DOOB, L. W., MILLER, N. E., MOWRER, O. H., & SEARS, R. R. *Frustration and aggression*. New Haven: Yale Univer. Press, 1939.

FREUD, S. *Beyond the pleasure principle*. London: International Psycho-Analytical Press, 1922.

GOUGH, H. G. *California Psychological Inventory: Manual*. Palo Alto: Consulting Psychologists Press, 1957.

GRINDER, R. E. Parental child rearing practices, guilt, and resistance to temptation of sixth grade children. *Amer. Psychologist*, 1960, **15,** 399. (Abstract)

HOLLINGSHEAD, A. B., & REDLICH, F. C. *Social class and mental illness*. New York: Wiley, 1958.

JOHNSON, ELIZABETH Z. Attitudes of children toward authority as projected in their doll play at two age levels. Unpublished doctoral dissertation, Harvard University, Graduate School of Education, 1951.

LEVIN, H., & SEARS, R. R. Identification with parents as a determinant of doll play aggression. *Child Develpm.*, 1956, **27,** 135–153.

MACCOBY, ELEANOR E. The taking of adult roles in middle childhood. *J. abnorm. soc. Psychol.*, 1961, **63,** 493–503.

MACCOBY, ELEANOR E., & GIBBS, PATRICIA K. Methods of child rearing in two social classes. In W. E. Martin & Celia B. Stendler (Eds.), *Reading in child development*. New York: Harcourt, Brace, 1954.

MCDOUGALL, W. *An introduction to social psychology*. London: Mathuen, 1908.

MILTON, G. A. A factor analytic study of child rearing behaviors. *Child Develpm.*, 1958, **29,** 381–392.

SEARS, PAULINE S. Doll play aggression in normal young children: Influence of sex, age, sibling status, father's absence. *Psychol. Monogr.*, 1951, **65**(6, Whole No. 323).

SEARS, PAULINE S. Child rearing factors as related to playing of sex-typed roles. *Amer. Psychologist*, 1953, **8,** 431. (Abstract)

SEARS, R. R. Effects of frustration and anxiety on fantasy aggression. *Amer. J. Orthopsychiat.*, 1951, **21,** 498–505.

SEARS, R. R. Transcultural variables and conceptual equivalence. In B. Kaplan (Ed.), *Studying personality cross-culturally*. Evanston: Row, Peterson, 1960.

SEARS, R. R., & HOVLAND, C. I. Experiments on motor conflict: II. Determination of mode of resolution by comparative strengths of conflicting responses. *J. exp. Psychol.*, 1941, **28,** 280–286.

SEARS, R. R., MACCOBY, ELEANOR E., & LEVIN, H. *Patterns of child rearing*. Evanston: Row, Peterson, 1957.

SEARS, R. R., WHITING, J. W. M., NOWLIS, V., & SEARS, PAULINE S. Some child rearing antecedents of aggression and dependency in young children. *Genet. psychol. Monogr.*, 1953, **47,** 135–234.

WARNER, W. L., MEEKER, M., & EELLS, K. *Social class in America*. Chicago: Science Research Associates, 1949.

WHITING, J. W. M., & CHILD, I. L. *Child training and personality: A cross-cultural study*. New Haven: Yale Univer. Press, 1953.

(Received November 14, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 63, No. 3, 493-503

THE TAKING OF ADULT ROLES IN MIDDLE CHILDHOOD[1]

ELEANOR E. MACCOBY
Stanford University

This report examines one aspect of "identification." It is one of a series of reports on a longitudinal study, the initial phase of which was reported in *Patterns of Child Rearing* (Sears, Maccoby, & Levin, 1957), in which the child training practices of the parents of 379 kindergarten-aged children in the greater Boston area were described. At that time, we also reported some relationships between early childhood parental practices and several personality characteristics of their 5-6 year old children. A follow-up study was done 6 years later, with those children from the original study whom we could locate at this later date. (A more detailed description of this study appears in Sears, 1961.) During the follow-up study, we did not obtain new information concerning the parents and their methods of rearing their children, but we did obtain fairly extensive information about the characteristics of the children, who were in the sixth grade at the time of our follow-up measurements. We can therefore report three kinds of findings: the interrelationships between several personality measurements taken at age 12; the interrelationships between certain aspects of child behavior at the two different ages we studied; the relationships between early childhood parental practices and the characteristics children display during middle childhood. The present report deals primarily with the first and third kinds of relationships.

During the process of early childhood interaction with his parents, the child, of course, learns to perform a host of actions that his parents deem desirable for a child, ranging from manual skills such as tying his shoes to temperamental matters such as controlling his temper. These kinds of behavior he learns in large part through guided trial and error, retaining primarily the behavior his parents reward, and eliminating what they do not approve of. There is another class of behavior which we believe the child learns, even though he seldom performs the behavior overtly and is therefore seldom directly reinforced for it. This is the behavior which forms part of the adult role performed by his parents in caring for him, but which would be inappropriate for him to perform toward them in return. Examples of adult role behavior would be setting rules, applying discipline, and giving medicine when the child is sick.

At first the child's learning of adult role behavior may be manifested overtly, as when a 2- or 3-year-old child is heard giving stern commands and admonitions to himself or his dolls, in a tone of voice patterned exactly after his parents when they say controlling phrases to him. But the child soon learns that he is not allowed to do or say certain things that he sees adults doing—he cannot scold adults, for example, nor start the car. We assume that although he ceases to do and say items of adult role behavior overtly, he does continue to practice them covertly, and continues to learn a great deal of the content of the behavior that characterizes the people with whom he interacts. A detailed description of how we believe this learning takes place has been given elsewhere (Maccoby, 1959). The sentences and other units of behavior thus learned are ready for elicitation when the appropriate situation occurs, and there may be a considerable lapse in time between the time they are learned and the time they are overtly manifested.

While all children presumably learn at least some elements of the adult role behavior that characterizes their parents, there should be

[1] From the Laboratories of Human Development of Stanford University and of Harvard University. This article reports a segment of a larger research project for which the research design and the collection of data were done in collaboration with J. W. M. Whiting, R. R. Sears, and Wesley Allinsmith. R. R. Sears has collaborated in analysis of the data. J. W. M. Whiting took an especially active part in the development of the concepts and theory which underlie the measurements and the hypotheses reported in this segment of the larger study.

The earlier stages of this research at Harvard were supported largely by a USPHS grant (M 461). The follow-up study was supported by an additional grant to Harvard (M 1468) and a mental health program grant to Stanford University by the Ford Foundation.

differences among children in the *amount* of adult role behavior they learn and the strength of the learning. As in all learning, the strength of a behavior tendency acquired through covert practice varies, we assume, with the amount of practice. Some children will frequently imagine themselves saying parent-like phrases and doing parent-like acts; others will seldom think of these things, and will therefore establish only weak tendencies to display parent-like actions later on. What are the conditions that govern the amount of a child's covert practice of adult roles?

We believe that several motives underlie such practice. The first is the child's desire to reproduce in fantasy events that have been reinforcing or associated with reinforcement. A parent's soothing, affectionate, and praising statements should be remembered by the child and rehearsed by him, especially in situations where the parent is not present so that the child must provide reassurance and social support for himself.

The second motive underlying a high level of covert practice of adult roles is based, we believe, on a high degree of parental control over the child. If the child cannot satisfy his needs without getting some mediating behavior from his parents (getting their permission to visit a friend, spend money, etc.), then his "vicarious trial and error" will necessarily involve rehearsing the kinds of controlling phrases his parents would say to him if he asked for their help in getting what he wanted. If he is not required to get parental permission for most of the steps he follows in pursuing his goals, then he will not have to take parental reactions into account in making his plans and will not engage in extensive covert practice of parental phrases and strictures.

We have suggested that there are two motives a child may have for practicing covertly the words and actions characteristic of his adult caretakers. One is to enjoy once more the satisfactions that come from rewarding, nurturant, anxiety reducing behavior directed toward him by his caretakers. The second is to anticipate the consequences of possible acts of his own, in order to work out good strategies for reaching his goals. A third motive will be mentioned here only briefly: to obtain vicarious satisfaction by pretending to be a person who is enjoying rewards that are denied to the child himself. The child's fantasy role playing of fictional characters has a large component of this motive, we think. The child likes to play that he *is* Davy Crockett or Superman, because then he can vicariously win the admiration of others and perform a great variety of feats of skill and strength. Presumably, the child can also get some vicarious satisfaction by pretending to be his parents as they interact with each other and the outside environment, since they have so much more power than he does and direct access to more sources of gratification. But for most of the discussion below, we will concentrate upon the first two motives, which have to do with the child's learning of those aspects of adult role behavior that are directed toward himself, the child, and are elicited by some feature of the child's own behavior.

As just mentioned, much adult role behavior is not appropriate for a child to display overtly vis-à-vis his parents. Indeed, he may be punished for displaying it. If he continues to practice it covertly, and thus establishes behavior potentials (habits) that are ready for elicitation, what will the situations be in which we may expect the behavior to be overtly manifested? Of course, when the child grows up and has children of his own, the actions which he has learned that are appropriate to the role of parent should emerge. But we expect that some elements of adult role behavior will emerge earlier, as the child interacts with his age-mates. Specifically, a child's adult role repertoire should be called out in relation to his age-mates whenever one of his age-mates provides the stimulus by acting out some aspect of a reciprocal child role. Thus if an age-mate solicits nurturance by making a dependent response, the child will make an adult role nurturant response, provided this response is well established through earlier covert role playing of a nurturant parent's role. Similarly, if a child sees his age-mate breaking a rule, this should elicit rule enforcement on his own part, provided again that his adult role learning has been strong.

The reader may have noted that although we used the term *identification* in the opening sentence of this article, we have not since

referred to this concept explicitly. We conceive of covert practice of adult role behavior as a basic process underlying the child's acquisition of behavior tendencies that are matched to the perceived behavior of a model. We are therefore talking about a process, or a product, that has frequently been labeled "identification." We have avoided use of the term because it has such a wide range of meanings in the existing literature. Also, we wanted to concentrate on both conditions of acquisition and conditions of elicitation of the model-matched behavior in question, and felt that this distinction was better handled with the terminology we have employed. Nevertheless our work took the concept of identification as its point of departure, and we consider that the phenomena we are studying do come under the heading of identification as the term is generally employed.

The present study concentrates upon the following issue: is there a tendency for children to employ, in their social interactions with age-mates, the same kinds of behavior their parents used toward them in similar situations? Can we discover conditions that determine which children will have taken on their parents' characteristic adult role behavior to a high degree and which will not? We have chosen, as the aspect of adult role behavior for study, the child's tendency to enforce rules when another child deviates. Our hypotheses are as follows:

1. When a child sees another child breaking a rule, he will be more likely to try to enforce the rule if his parents were strict rule enforcers during his early childhood. To test this hypothesis we employ two measures of the parents' rule enforcing behavior: their restrictiveness and their punitiveness.

2. The match between parental behavior and that of the child with respect to rule enforcing will be closest in those families where the parents were highly nurturant.

METHOD

Sample

Our measures of sixth graders were taken in the spring of 1958, in selected schools in the greater Boston area. The subjects were a group of children some of whom had been included previously in the study reported in *Patterns of Child Rearing* (Sears et al., 1957). The design and sampling procedures for the earlier study are described in some detail in that book. Similar details for the follow-up study appear in Sears (1961). Most of the children in our original sample had, by the spring of 1958, reached the sixth grade. We administered most of our measuring instruments to all the children in the sixth grades in the schools where we did our work. We obtained measures for the majority of our scales on 525 children, of whom 165 had been in the original sample 6 years earlier. In the following report, our analysis is based primarily upon the children who were in the original sample, although occasionally findings from the larger sample can be used.

Measures

The primary scale to be considered in the following report is Enforcing Rules. The measuring instrument that included the items for this scale also contained items for a number of other scales, which we describe briefly since we will report their interrelationships with the Rule Enforcement scale.

There were seven scales scored from our so-called "Role Taking" questionnaire.[2] This questionnaire was administered during a regular classroom hour by a research assistant. Most of the items were two or three-choice items, and each child checked his own choices after a silent reading of the items. For each scale, an item analysis was carried out on 100 cases from our final sample, not including any of the children in our original study. We correlated each item with total scores on the scale of which the item was intended to form a part, and on the basis of this analysis, discarded a few items that proved unreliable on the basis of internal consistency. For the scales where item analysis revealed generally poor scale reliability, we examined the inter-item correlations to see whether we could detect coherent groups of items that could be used to form one or more subscales. In Table 1 appears a description of each scale finally employed, giving an example of an item, the number of original items, and the number dropped during item analysis. In each case, the example given is taken from the boy's form. For the girls' form of the questionnaire, the items were changed to refer to another girl with whom the subject was interacting.

The Adult-Child Role Choice scale had two classes of items. The first involved the child's choosing between adult-like functions and child-like functions when both could not be chosen. The second involved situations in which another child deviated and was caught by an adult, and the question was which of the two (adult or deviating child) the subject "sided with," or empathized with. While we feared these diveres items might not turn out to form a unidimensional scale, item analysis did reveal positive relationships

[2] The seven scales and a table presenting their intercorrelations have been deposited with the American Documentation Institute. Order Document No. 6869 from AD1 Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

TABLE 1
SCALES SCORED FROM ROLE TAKING QUESTIONNAIRE

Scale	Example of an item	Original number of items	Number retained
Rule enforcement	You are on your way to school. A boy in your class is walking in front of you. He drops a bottle on the sidewalk, and it breaks in many pieces. The boy walks on. What would you do? ____Say nothing; it's his business, not mine. ____Tell the boy to pick up the pieces.	6	6
Acceptance of rule enforcement	By accident, you knocked down a bunch of coats in the coat room. Another boy said to you, "Hey, you better pick them up." What would you say? ____Don't worry, I was just going to. ____O. K. Thanks for telling me.	6	6
Give nurturance	A boy in your class asks you to help him with his homework. You think this boy could do his homework by himself if he really tried. What do you do? ____Tell him I haven't time. ____Tell him to do his own homework and stop bothering me. ____Help him.	7	5
Accept material help	You and your friend have each bought a bag of candy. You finish yours first, and you are still hungry. He offers you one of the pieces of candy he has left. What do you do? ____Thank him and take one. ____Say, "No, thanks."	5	4
Accept comfort	Suppose you fall down and hurt yourself. You aren't hurt badly, but you've skinned your arm and leg and you feel shaky. One of the other boys offers to stay with you till you feel better. Would you rather someone would stay with you, or would you rather they'd leave you alone? ____I'd like someone to stay with me. ____I'd rather they'd leave me alone.	4	4
Beg for return of nurturance	Suppose there was a boy who had been quite a good friend of yours. He starts acting less friendly. He doesn't wait for you after school as often as he used to. What do you do? ____Tell him I think he's being unfair. ____Ask him to wait after school the way he used to. ____Don't do anything—it will blow over.	5	5
Adult-child role choice	1. You are at a school movie. You can either sit with your friends and watch the movie, or be an usher and show people to their seats. Which would you rather do? ____Sit with my friends and watch the movie. ____Be an usher and show people to their seats. 2. In your school, it's against the rules to slide down the banisters. One day, a boy you know slides down, and at the bottom he is met by a teacher. You are watching. How do you feel? ____Sorry for him, because he got caught. ____Not sorry for him; it's his own fault for breaking the rules.	9	9

between the two kinds of items, and between individual items and total score. Consequently, we did not subdivide our scores on Adult-Child Role Choice.

In the present report, we consider the relationship of rule enforcement not only to the scales just described, but to a set of scales on aggressive tendencies. These scales are described in a separate publication (Sears, 1961).

During the follow-up study, we interviewed the classroom teacher about each of the children in her room who had been included in our original study 6 years earlier. We asked about a number of aspects of the child's behavior in the classroom, including his relationship to the other children in the class as well as to the teacher. In the present report, we present a few findings from the teacher interview as they relate to rule enforcing tendencies.

RESULTS

Relation of Adult Role Functions to Other Personality Traits

First we consider the meaning of rule enforcement in the sense of its interrelations with other characteristics of the children.

TABLE 2

CORRELATIONS BETWEEN SCORES ON RULE ENFORCEMENT SCALE AND OTHER PERSONALITY MEASURES

(Total sample)

Measure	Boys ($N = 261$)	Girls ($N = 264$)
Accept rule enforcement	.24	.06
Accept comfort	.19	.11
Accept material help	−.05	.06
Beg for return of nurturance	.13	.16
Adult-child role choice	.34	.25
Give nurturance	.10	−.08
Self-criticism	−.02	−.10
Aggression anxiety	.18	.17
Projected aggression	.06	.00
Self-aggression	.02	.02
Prosocial aggression	.24	.08
Antisocial aggression	−.18	−.12

TABLE 3

RELATIONSHIP BETWEEN TEACHER-INTERVIEW REPORTS AND CHILD'S SCORES ON RULE ENFORCEMENT

(Follow-up sample only)

	Boys rule enforcement score		Girls rule enforcement score	
	Low (0–3)	High (4–6)	Low (0–3)	High (4–6)
Teacher's report, percentage who:				
Are moderately or markedly aggressive toward peers	45%**	13%**	18%	25%
Misbehave when teacher is out of room	35%	31%	10%**	35%**
Are relatively highly dependent on teacher	45%	33%	28%	36%
Are highly motivated to do their school work	33%*	60%*	79%*	57%*
Number of cases	32	45	50	37

* Significantly different at <.05 level.
** Significantly different at <.01 level.

Table 2 shows the several scales derived from the role taking and aggression questionnaires, and their intercorrelations with "enforcing rules."

Table 2 shows that, for both sexes, the correlations between "enforcing rules" and the other personality traits we measured are low. But quite a few are significant: a correlation of .17 is significant at the .01 level with a sample of this size. The table, then, shows the following small but significant relationships: first, children of both sexes who select the rule enforcing alternative to our questions tend to be "well socialized" in other respects. When faced with a choice between adult versus child-like forms of behavior, the rule enforcing child chooses the adult forms more often than other children. Furthermore, rule enforcement for both sexes is positively related to aggression anxiety.

The meaning of rule enforcement appears to be different in some respects for the two sexes. For boys, requiring other boys to obey rules tends to be accompanied by a willingness to accept rule enforcement from other boys when one has deviated oneself. This is the reciprocity which Piaget (1948) believed was a central element in mature moral development. For girls, we do not find a correlation between the tendency to enforce rules on others and the willingness to accept rule enforcement in turn. Piaget noted that moral development among girls did not appear to be as dependent upon the development of a reciprocal relationship with peers as was the case with boys, and our finding confirms his observation.

The sexes also differ in the relationship of "rule enforcement" to characteristics that relate to the handling of aggression. For boys, rule enforcement again shows itself as a symptom of successful socialization in that boys who give many rule enforcing responses are low in antisocial aggression and high in prosocial aggression, as measured by our self-administered scale. For girls, the relationships between these traits are in the same direction but weaker and not statistically significant.

We see some confirmation of the trends we have reported in Table 3 when we study the teacher interviews and relate the characteristics the teachers report to the "rule enforcing" scores derived from the self-administered scales. We see that, for boys, high scores on both these scales are associated with "good" behavior in the classroom; that is, high-scoring boys are reported by the teachers to be unlikely, as compared with other boys, to start fights with their peers; and the teachers report that they work especially hard at their classroom assignments. None of these relationships holds for girls. In fact, the girls who score high on rule enforcing tendencies are lower than other girls in their interest in their school work, and more likely to misbehave when the teacher leaves the room.

In view of the fact that the tendency to enforce rules forms part of somewhat different constellations of traits for the two sexes, we must be prepared for the possibility that different aspects of early childhood training may predict it for each sex. Nevertheless, we proceed similarly for the two sexes with an analysis that follows from the theoretical considerations outlined at the beginning of this report.

Antecedents of Rule Enforcement

We predicted earlier that if parents had been strict rule enforcers toward a child during his early years, the child would in his turn be a strict rule enforcer with his peers in middle childhood. For our first test of this hypothesis, we employ the "summary permissiveness" scale derived from our early mother interviews. This scale was based on a factor analysis of a large number of mother-interview scales, and is composed of a sum of the following scores: sex permissiveness, permissiveness for aggression toward parents, permissiveness for dependency, and low restrictions on the child's use of house and furniture. High scores on this scale mean that the mother was permissive (nonrestrictive) in all or most of these spheres of behavior. Our prediction is that if the mother was *restrictive* (scored on the low end of our scale), her child should have relatively strong tendencies to enforce rules upon other children.

As may be seen in Table 4, the prediction is borne out for boys, but not for girls. Boys whose mothers were restrictive in early childhood give significantly more rule enforcing responses at age 11 than boys whose mothers were permissive during the child's early years. For girls, the aspect of child rearing that predicts the number of the girl's rule enforcing responses is not how restrictive the parents were in early childhood, but how punitive they were. Punitiveness is again measured by a summary scale, based on our factor analysis of the mother interviews. We added scores on the three variables that proved to have the highest loadings on our punitiveness factor: punishment for aggression toward parents, amount of physical punishment, and demands for the child to be aggressive toward other children. We also added the scores on the mother's irritation over the child's dependency.

TABLE 4

RELATIONSHIP OF EARLY CHILDHOOD SOCIALIZATION PRACTICES TO CHILDREN'S SCORES ON RULE ENFORCING

(Follow-up sample only)

	Mean score: rule enforcement	
	Boys	Girls
Mother's permissiveness:		
Low (1–4)	4.28	3.26
	$t = 2.64$	
High (5–9)	3.39	3.07
Mother's punitiveness:		
Low (1–5)	3.76	2.61
		$t = 3.3$
High (6–9)	4.08	3.61
Correlation with rule enforcement scores[a]		
Mother's permissiveness	−.25	−.10
Mother's punitiveness	.03	.23
Number of cases	78	87

[a] With a sample of this size, a correlation of .22 is significant at the .05 level, and one of .28 at the .01 level.

This scale was not included in the factor analysis, but in intercorrelations done subsequently, it proved to be part of the punitiveness cluster. As Table 4 shows, when mothers were punitive toward their daughters in early childhood, these daughters tended, several years later, to be high rule enforcers with peers who deviated.

Boys: Conditions Predicting a Close Match to Parental Behavior

Let us concentrate now on the relationship between maternal restrictiveness and rule enforcing among boys. Our original hypothesis was that children should resemble their parents in the rule enforcing tendencies they displayed. But we further predicted that some boys should resemble their parents more than others. We said that a condition governing a close match should be high parental warmth. Our reason for expecting that parents who were warm and nurturant toward the children would have children more similar to themselves than less warm parents was that the child's covert role practice of the parents' behavior should be increased if the child could give himself pleasure by imagining things that his parents had said and done toward him. We are *not* saying that warm mothers should have rule enforcing children. We are saying that if

TABLE 5

RELATIONSHIP BETWEEN PARENTAL RESTRICTIVENESS AND THE CHILD'S RULE ENFORCING (Boys)

	Percentage of boys "high" on rule enforcing	*N*	Difference between percentages	Difference between differences
Mother relatively warm:				
Mother restrictive	81%	16	43% ($p < .01$)	32% ($p < .09$)
Mother permissive	38%	21		
Mother relatively cold:				
Mother restrictive	64%	22	11% (*ns*)	
Mother permissive	53%	17		

a mother is warm, her level of rule enforcing should be closely reflected in that of her child, so that if she is warm and restrictive, the child should be high in rule enforcing, and if she is warm and permissive, her child should be low in rule enforcing. If the mother is cold, on the other hand, we do not expect her degree of restrictiveness to make as much difference in the rule enforcing tendencies of her child. We see that our prediction is at least partially borne out, for (among boys) the relationship between parental restrictiveness and the child's rule enforcing is indeed somewhat closer when the mother is warm (see Table 5).

Within this group, there is a significant relationship between maternal restrictiveness and the son's rule enforcing, despite the small number of cases per cell. When the mother is cold, the relationship is small and not significant. However, analysis of the second-order differences[3] does not show a significant effect of maternal warmth on the size of the relationship between restrictiveness and rule enforcing, so the interaction we predicted is suggested but not demonstrated.

[3] Here and in the instances which follow, the significance of second-order differences has been computed by the following procedure: The χ^2 value for each of the first-order differences was computed, and from this, the standard error of the difference was obtained (from the formula: $\sqrt{\chi^2} = D/\delta d$ for fourfold tables). A standard error for the difference between differences was then obtained from the formula:

$$\delta_{d-d} = \sqrt{\delta_{D_1}^2 + \delta_{D_2}^2}$$

and a critical ratio computed. One-tailed tests of significance have been employed in these second-order analyses, since the direction of differences was predicted.

There are, of course, other conditions than parental nurturance that should cause the child to have his attention fixed on his parents and thus learn their characteristic responses. We can make the prediction that whatever the origin of dependency in a particular child, the children who were highly dependent during the preschool years should have acquired more of their parents' characteristics and therefore show a closer match in their rule enforcing tendencies at a later time. This prediction is moderately well borne out (see Table 6).

The analysis of second-order differences shows that there is a tendency (significant at $p = < .05$) for the boy's early-childhood level of dependency to make a difference in the closeness with which his rule enforcing behavior is matched to that of his parents, and the nature of the relationship is as predicted.

In our earlier study (Sears et al., 1957, p. 168), we did not find any direct relationship between the mother's warmth and the child's dependency; so we can regard the two foregoing sets of tabulations as tests of two independent hypotheses.

We see that it is the child's dependency *on his parents* that is associated with a close match between their behavior and his. We should not expect the same results if we consider a child's dependency on his peers. That is, the child who is dependent on emotional support from his age-mates at school, in the sense of being anxious to have their approval and anxious to regain their approval when it is withdrawn, will have taken new models for his behavior. We would expect that the child who shows the closest match to his parents' behavior

TABLE 6

RELATIONSHIP BETWEEN EARLY DEPENDENCY AND THE CHILD'S RULE ENFORCING (Boys)

	Percentage of boys "high" on rule enforcing	*N*	Difference between percentages	Difference between differences
Boys relatively dependent at preschool age, and:				
Mother restrictive	70%	23	45% ($p < .01$)	40% ($p < .05$)
Mother permissive	25%	20		
Boys relatively nondependent at preschool age, and:				
Mother restrictive	73%	15	5% (*ns*)	
Mother permissive	68%	19		

TABLE 7

RELATIONSHIP BETWEEN NURTURANCE BEGGING AND RULE ENFORCING (Boys)

	Percentage of boys "high" on rule enforcing	*N*	Difference between percentages	Difference between differences
Boys low on begging for return of nurturance from peers, and:				
Mother restrictive	74%	19	41% ($p < .01$)	34% ($p = .06$)
Mother permissive	33%	21		
Boys high on begging for return of nurturance from peers, and:				
Mother restrictive	68%	19	7% (*ns*)	
Mother permissive	61%	18		

should be the one who was both dependent upon them during early childhood, and *not* especially dependent upon peers currently. If he has shifted from parent dependency to peer dependency, we expect that although he will have learned earlier to act as his parents did, these tendencies will have been supplanted by new learning based upon his new focus for dependency. To discover whether current dependency on peers does, indeed, weaken the match between parent and child behavior, we employed our measure of the child's tendency to beg for the return of nurturance when peers have withdrawn it (see Table 7). We see that the relationship between the mother's strictness and her son's tendency to enforce rules on peers is weaker when the boy is dependent on his peers in the sense of dependency we have used. This difference in the size of the relationship under the two conditions of early dependency is strongly suggested by our data, although a chance hypothesis cannot be rejected with a p value of .06.

We have seen, then, that a boy's tendency to act toward peers as his parents acted toward him is greater when he was highly dependent upon them. It further follows from our theory that the similarity in behavior between the two generations should be greater when the parents exercised close control over the children. This is a difficult hypothesis to test if the dependent measure we are using is the child's restrictiveness toward peers. The only way we can approach the question is through the standard deviation of the child's responses. We predict that when the parent is restrictive, the child should be restrictive (rule enforcing) toward peers, and that there should be relatively little variability in rule enforcing among the children of restrictive parents. But when the parents are nonrestrictive, our prediction is that the children will have failed to copy

them, so that the child may or may not be a rule enforcer. His behavior should simply be unrelated to that of his parents. What we are predicting, essentially, is a scatterplot with an empty quadrant. That the relationship between the parent's restrictiveness and the son's rule enforcing is somewhat of this nature is indicated by the following facts: when the parents are restrictive, the sons are not only higher in rule enforcing, but the standard deviation of their scores is relatively low ($\delta = 1.69$); for boys whose parents were permissive, their mean score on rule enforcing is lower, as we saw in Table 4, but the dispersion of their scores is somewhat, but not significantly, greater ($\delta = 2.55$). This suggests that when a parent is restrictive, his sons will be restrictive. When the parent is permissive, the son tends to be either permissive or restrictive toward peers. His behavior tendencies in this sphere will, we suggest, have been acquired in some other way than by copying the behavior of his parents.

We had a subsidiary hypothesis, not previously mentioned, that the child's behavior would be more like that of his parents if authority in the household had been primarily exercised by the same-sexed parent. Our rationale for this prediction was that, for reasons already given, we would expect a child to acquire more of his action tendencies from the more powerful of the two parents; if these action tendencies were acquired from the same-sexed parent, there would be little conflict between them and other things the child has learned about the appropriate sex-typing of behavior; so they should be more likely to find their way into the child's overt behavior at a later time than would be the case with behavior tendencies acquired from an opposite-sexed parent. Our prediction on this score is *not* borne out. We classified the families of the children we studied according to the identity of the parent who was stricter and who was more likely to be the one to administer discipline when it was required. There was no greater matching between the boy's rule enforcing and the parent's restrictiveness when his father was the authority figure in the household than when the mother played this role.

Girls: Conditions Related to the Closeness of Matching to Parental Behavior

We have been pursuing the conditions that govern the closeness of the relationship between a boy's rule enforcing and his parents' permissiveness or restrictiveness. Let us now turn to the girls, and see whether any similar regularities can be discovered. We have already seen that it is punitiveness rather than restrictiveness on the part of the parents that is associated with a girl's being a strict rule enforcer with her peers. Why there is this sex difference in the home conditions associated with a child's rule enforcing tendencies we do not know. We suspect, however, that in view of the difference in antecedent conditions underlying rule enforcement, and the difference in the pattern of other personality traits associated with it, rule enforcement is a qualitatively somewhat different thing when girls do it than when boys do it. For girls, it probably has more aggressive overtones. In any case, we assume for the ensuing discussion that a girl who is high on rule enforcing is displaying toward her peers the same kind of behavior her mother displayed toward her when the mother was punitive during the girl's early childhood. Having already shown that there is an association between maternal punitiveness and the daughter's later rule enforcing tendencies, we now explore the conditions that govern the closeness with which the daughter's behavior is related to her mother's. That is, as we did with boys, we now try to discover what other conditions (either in the child training practices prevailing in the home, or in the personality traits developed by the child) increase the correlation between mother's and daughter's behavior, and what conditions reduce it.

When we subdivide the girls according to whether their mothers were predominantly warm or cold toward them in early childhood, we do not find that this makes any difference in the size of the relationship between parental punitiveness and the girl's rule enforcing tendencies. The girls' earlier dependency, however, does make a difference. Just as we found with boys, the girl's behavior toward her peers is a closer reflection of her mother's earlier behavior toward her, if the girl was relatively dependent as a young child (see

TABLE 8

RELATIONSHIP BETWEEN DEPENDENCY AND THE CHILD'S RULE ENFORCING (Girls)

	Percentage of girls "high" in rule enforcing	*N*	Difference between percentages	Difference between differences
Girls relatively dependent at preschool age, and:				
Mother punitive	76%	25	65% ($p < .001$)	57% ($p < .001$)
Mother not punitive	11%	18		
Girls relatively nondependent at preschool age, and:				
Mother punitive	40%	22	8% (*ns*)	
Mother not punitive	32%	19		

TABLE 9

RELATIONSHIP BETWEEN NURTURANCE BEGGING AND RULE ENFORCING (Girls)

	Percentage of girls "high" on rule enforcing	*N*	Difference between percentages	Difference between differences
Girls low on begging for return of nurturance from peers, and:				
Mother punitive	71%	17	58% ($p < .001$)	35% ($p < .06$)
Mother not punitive	13%	15		
Girls high on begging for return of nurturance from peers, and:				
Mother punitive	50%	32	23% ($p < .11$)	
Mother not punitive	27%	22		

Table 8). And again, as we did with boys, we inquired whether the girl's behavior toward her peers is more closely matched to that of her mother if the girl is not currently especially dependent on her peers. We find that the relationship between maternal punitiveness and the daughter's tendency to enforce rules toward peers is greater if the girl is not dependent on her peers—at least as we have measured this dependency through our Begging for Return of Nurturance scale. The second order difference reaches only the .06 level of significance, however (see Table 9).

It should be noted that the over-all proportion of high rule enforcers is about the same for the girls who are high on current dependency toward peers as it is for those who are low, disregarding the mother's punitiveness. In fact, there is a small *positive* correlation in the sample as a whole between rule enforcing and begging for return of nurturance from peers. Therefore, our findings do *not* mean that when a child is anxious to keep the good opinion of age-mates he or she will avoid enforcing rules, for fear of alienating his peers by a show of seeming bossiness. Instead, the foregoing table (and the parallel one for boys) means that among these peer-dependent children, the mother's punitiveness is not the determining factor in whether they will be rule enforcers, while among the children who are more independent of their peers' approval, rule enforcing tends to be matched to what their parents would do in similar situations.

For girls, too, we inquired whether the mother-daughter similarity in rule enforcing is affected by whether the mother is the major agent of discipline during the child's early years. We find, to our surprise, that among girls the correlation between parental punitiveness and the child's rule enforcing is

somewhat (though not significantly) closer when the *father* is the agent of discipline. The hypothesis, then, that having a same-sexed agent of discipline produces greater matching of the child's behavior to that of the parent is clearly not sustained for either sex.

SUMMARY

This is one of a series of reports on a follow-up study done with some of the families originally described in *Patterns of Child Rearing* (Sears et al., 1957). The present report deals with one aspect of moral development in middle childhood (sixth grade): namely, the child's tendency to enforce rules when other children break them. Scores on a scale measuring this tendency were correlated with measures, taken concurrently, of other personality traits. Then the subjects' rule enforcing tendencies were related to certain aspects of the child rearing practices employed by their parents during the children's preschool years. The major findings are as follows:

1. Rule enforcing is part of a somewhat different constellation of personality traits for the two sexes. For example, among boys it tends to be accompanied by a reciprocal tendency to accept rule enforcement from peers when the subject himself has deviated, while among girls this reciprocity is not found.

2. Rule enforcing boys have parents who were fairly strict rule enforcers during the child's earlier years, in the sense that they were restrictive (nonpermissive) about a variety of childish impulsive behavior. The match between the son's rule enforcing tendencies and the earlier behavior of his parents is significantly closer when the child was relatively dependent during early childhood, and tends to be closer when (*a*) the parents were relatively warm, and (*b*) the child is *not* currently especially dependent upon peers in middle childhood.

3. Rule enforcing girls tend to have had relatively punitive parents during early childhood. The match between parental punitiveness and the daughters' rule enforcing tendencies is significantly greater when the child was relatively dependent upon the parents in early childhood, and tends to be greater if the child was not strongly dependent upon peers at the time of the follow-up study. The findings are interpreted as supporting the view that the child acquires a set of adult-like behavior tendencies during early childhood, through covert practice of his parents' actions, and that these tendencies find their way into overt expression during interaction with peers at a later time when peers provide the necessary stimulating conditions.

REFERENCES

MACCOBY, ELEANOR E. Role-taking in childhood and its consequences for social learning. *Child Develpm.*, 1959, **30**, 239–252.

PIAGET, J. *The moral judgment of the child*. Glencoe, Ill.: Free Press, 1948.

SEARS, R. R. Relation of early socialization experiences to aggression in middle childhood. *J. abnorm. soc. Psychol.*, 1961, **63**, 466–492.

SEARS, R. R., MACCOBY, ELEANOR E., & LEVIN, H. *Patterns of child rearing*. Evanston Ill.: Row, Peterson, 1957.

(Received November 14, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 63, No. 3, 504-510

GROUP THERAPY OF PHOBIC DISORDERS BY SYSTEMATIC DESENSITIZATION[1]

ARNOLD A. LAZARUS

Psychiatric Unit, Witwatersrand University Medical School

THE increasing demands for psychological and psychiatric services dictate the need for effective short-term therapeutic techniques and the extension of the existing services. Consequently, group techniques have grown in clinical stature, and the past decade has witnessed the development of numerous divergent procedures. A most promising variety of short-term therapy is Wolpe's (1958) system of "reciprocal inhibition," by which he achieved the recovery of 188 out of 210 neurotic cases in an average of 34.8 sessions.

A double economy can be achieved by combining the advantages of Wolpe's (1958) expedient clinical procedures with the additional time- and effort-saving properties of group therapy. This paper describes the adaptation of Wolpe's most important therapeutic procedure—the technique of systematic desensitization based on relaxation—to the group treatment of phobic disorders. In addition, the therapeutic effects of group desensitization were compared with more conventional forms of interpretive group psychotherapy on matched pairs of phobic subjects.

METHOD

General Procedure

The sample consisted of 35 middle-class urban white South Africans who were handicapped by phobic disorders.[2] Social class membership was defined in terms of education, vocation, and income. There were 7 university graduates, 16 matriculants, and 12 patients with at least 3 years of secondary schooling. Apart from 3 professional women, the majority of female patients were housewives whose husbands' average earnings were the equivalent of $550 a month. The mean income for the rest of the group was approximately $600 a month. In all, there were 12 men and 23 women, the mean age being 33.2 years with a standard deviation of 9.87.

The entire group included 11 acrophobics, 15 claustrophobics, 5 impotent men (treated as suffering from sexual phobia), and a mixed group of 4 phobic patients. The latter comprised a girl with a fear of sharp objects, a man with a fear of physical violence, a woman who was afraid to be a passenger in a moving vehicle, and a woman with a phobia for dogs.

The basic experimental design was to compare group desensitization therapy with more conventional methods of group treatment (or therapy based on "group dynamics"). The group desensitization technique consisted of systematically counterposing by relaxation graded lists of anxiety evoking stimuli which the separate groups of patients were asked to imagine.

The efficacy of group desensitization was first compared with group interpretation. The same therapist (the investigator) conducted all the therapeutic groups.

The initial comparison was made on a group of five acrophobic patients, two of whom received desensitization therapy, and three who were treated by interpretive group procedures.

Throughout the experiment, pairs of phobic patients were matched in terms of sex, age (within a 4-year range), and the nature and objective severity of the phobic disorders. A coin was tossed to decide whether a given member of each matched pair would be treated by desensitization therapy or by group interpretation. Extra (unmatched) individuals were always placed in the interpretive groups because it is generally agreed that these groups require a minimum of three members (Corsini, 1957).

When, after six sessions, treatment with the acrophobic groups was well underway, the next group (five impotent men) was selected and similarly subdivided into two additional groups, treated by desensitization and interpretation, respectively.

Three months later, the group of six claustrophobic patients was selected and equally subdivided to form a third separate desensitization-interpretation comparison.

Thus, at the end of 6 months, a total of seven patients had received group desensitization and nine had been treated by group interpretation.

Seven months later, additional acrophobic and claustrophobic patients were obtained in order to investigate the effects of relaxation per se. It was hypothesized that individuals who received training in relaxation at the end of each interpretive session would show a greater diminution of phobic reactions than those patients who had been treated solely by group

[1] This paper is an outline of the experimental section of a thesis entitled "New Group Techniques in the Treatment of Phobic Conditions," which was accepted by the University of the Witwatersrand in December 1960, for the degree of Doctor of Philosophy.

[2] The sample was not drawn from psychiatric hospitals or institutions as it was felt that extraneous variables would be introduced. Since the rules of the South African Psychological Association forbid registered psychologists to advertise in the press, the patients were obtained with the generous aid of friends and colleagues who made announcements at lectures and contacted their own associates.

interpretation. The suggestion that interpretation-plus-relaxation might be more effective than desensitization was also tested.

Accordingly, an additional six acrophobic patients were equally divided into two matched groups, the one receiving desensitization therapy and the other receiving interpretation-plus-relaxation. The latter group was trained in an accelerated version of Jacobson's (1938) progressive relaxation for about 15 minutes at the end of each interpretive therapeutic discussion.

A few weeks later, an additional nine claustrophobic patients were similarly subdivided. Group desensitization was administered to four patients and group interpretation-plus-relaxation was applied to five.

Finally, the mixed phobic group was treated by desensitization in order to determine whether desensitization could be successfully applied to a heterogeneous phobic group.

Thus, group desensitization was applied to 18 patients; group interpretation was applied to 9 patients, and 8 patients were treated by group interpretation-plus-relaxation.

Selection of Phobic Patients

Although there were numerous volunteers for inclusion in the investigation, only those people whose phobias imposed a severe limitation on their social mobility, jeopardized their interpersonal relationships, or hindered their constructive abilities were admitted to the therapeutic groups.

Several people who were greatly handicapped by phobic disorders were excluded because they had received previous psychiatric treatment, ranging from psychoanalysis to electroconvulsive therapy. These people were given individual treatment in order to avoid ambiguity concerning the effects of the therapeutic groups.

The character and severity of the phobias were assessed in the following manner: Patients reporting acrophobic symptoms were privately and individually required to climb a metal fire escape. The experimenter climbed the stairs directly behind the patients and urged them to see how high they could climb. Few of the patients were able to proceed higher than the first landing (approximately 15–20 feet from ground level). The patients who were admitted to the acrophobic groups were all able to achieve a pretherapeutic height of between 15 and 25 feet.

Similarly, patients with claustrophobic traits were admitted individually into a well-ventilated cubicle with large French windows which opened onto a balcony. The patient sat facing the open windows. To the left of the patient was a movable screen which could be pushed as far as the centre of the cubicle, thus, creating a sensation of space constriction. The patients were told that the experimenter would first shut the French windows and then proceed to push the screen towards the centre of the room. They were urged to remain in the cubicle for as long as possible and to reopen the windows only when they felt that the need for air had become unbearable. Most of the patients showed visible signs of discomfort as soon as the windows were shut, and no one was able to tolerate the screen at a distance of less than 20 inches.

Detailed information regarding the purpose of the investigation was withheld from the patients to avoid possible prejudice to the results. They were merely informed that the experimenter was conducting research into the alleviation of phobic disorders by group methods.

Apart from the initial screening procedures, individual contact with the patients was avoided in order to exclude the influence of any additional therapeutic factors. It is thought, for instance, that history taking and psychometric investigations may in themselves be therapeutic. In order to determine the value of group therapy per se, it was considered necessary to eliminate as many of these extraneous variables as possible. Attention was deliberately focused, therefore, on the specific techniques under investigation, avoiding the use of any supplementary measures which might facilitate therapeutic progress. For instance, in clinical practice it is customary to precede the application of systematic desensitization by a brief outline of the theoretical rationale behind the technique. Since it could be argued that this practice has a direct bearing on the results, the patients were directly desensitized without any preliminary explanation.

Group Desensitization

Anxiety hierarchies (graded lists of stimuli to which the patients reacted with unadaptive anxiety) were constructed (Wolpe, 1958). In preparing these hierarchies, the experimenter extracted common elements from remarks which individual patients wrote on the questionnaires they filled out, and a group hierarchy was constructed. It must be emphasized that the hierarchical situations were imaginary ones, listed on paper and presented only symbolically to the patients.

The acrophobic group hierarchy, for example, consisted of the following situations: looking down from a very high building, seeing films taken from an airplane, looking down from a height of 80–100 feet, looking down a well, sitting high up on a grandstand during a football game, looking down from a 60-foot balcony, sitting on a narrow ledge at a height of 60 feet with a safety net a few feet away, looking down from a height of about 55 feet, seeing someone jump from a 50-foot diving board, sitting on a wide ledge about 35 feet from the ground, looking down from a height of approximately 20 feet, looking down from a height of about 10 feet.

The claustrophobic group hierarchy consisted of 16 situations ranging from "sitting in a large and airy room with all the windows open," to "sitting in front of an open fire in a small room with the doors and windows shut." The group hierarchy applied to the impotent men contained 10 items referring progressively intimate sexual situations requiring increasing amounts of initiative.

The first therapeutic session was devoted entirely to training the patients in intensive muscular relaxation. At the end of the session, the patients were instructed to practise specific relaxation exercises for about 15 minutes morning and night.

The second session was held 3 days later, when further training in relaxation was provided. Towards the end of this session, desensitization commenced

with the presentation (in imagination) of the two weakest items of the relevant anxiety hierarchy. The acrophobic patients, for example, were first told to picture themselves looking out of a window about 10 feet from the level of the street. It was impressed upon them that if any scene proved upsetting or disturbing, they were to indicate this by raising their left hand. When any patient signaled in this manner, the scene was "withdrawn" immediately.

When the two least disturbing items in the relevant anxiety hierarchy had been presented, each of the patients was asked to report on the clarity of the imagined scenes and their accompanying levels of disturbance. The second session ended after the patients had been told to practise relaxation twice daily for periods of about 10 minutes.

The subsequent desensitization sessions followed a set pattern. The therapist named the various muscle groups to be relaxed. When a deep level of relaxation was reached, the patients were presented with successive items from the hierarchical series. The desensitization procedure was conducted at the pace of the "slowest" (i.e., most anxious) subject.

The third session was terminated only when all the patients were able to tolerate an exposure of about 10 seconds to the first three items on the hierarchy without signaling anxiety. Thereafter, new items were introduced only when a 10-second tolerance to the preceding item had been achieved. It took several sessions before the entire group was able to visualize a given item for as long as 10 seconds without one or another member's signaling some disturbance.

Apart from occasional restlessness in those who were ready for more "difficult" anxiety items but who were constantly re-exposed to stimuli which they had long since mastered, no harm seemed to ensue from proceeding at a pace that was obviously too slow for part of the group. On the other hand, experience has shown that too rapid a pace can prove extremely antitherapeutic and lead to increased levels of anxiety.

Therapy in the desensitization groups was terminated when the final item on the hierarchy was tolerated by the patients for 10 seconds without signaling. Patients often reported a marked amelioration of their phobic responses when the anxiety hierarchies were only half completed. It was insisted that each member would nevertheless have to undergo desensitization of the entire hierarchy in order to consolidate and reinforce their therapeutic gains.

The treatment of the claustrophobic groups was conducted out of doors. The patients in the mixed phobic group were handed the items of their relevant anxiety hierarchies on slips of paper. Here, the relaxation procedure adopted was as previously outlined, but instead of describing the items, the therapist handed a typewritten anxiety scene to each group member and instructed him to read the description of the scene, to close his eyes and to try to imagine the situation with tranquility. The patients were instructed to signal in the usual manner when a given situation became disturbing, and then immediately to stop imagining the scene and to continue relaxing. After about 10 seconds, all the patients were told to stop picturing the scene and remain relaxed. Those who had successfully imagined their item without undue disturbance were then handed a new anxiety situation. In this manner, each group member was able to proceed at his own pace. No more than two successive items were presented at any one session.

Group Interpretation

The approach used in the interpretive groups was a form of insight therapy with re-educative goals (Wolberg, 1954). Leadership was basically democratic, and the therapist's primary role was that of a participant observer. The groups passed through two phases: First, there was an introductory period during which the group situation was structured with the emphasis on a free and permissive emotional atmosphere. Feelings of initial tension and reticence were dealt with by open discussion, emphasizing group tolerance and acceptance and clarifying numerous misconceptions. Second, descriptions of phobic symptoms preceded intensive discussions which focused attention on emotions and on current interpersonal relationships. The emphasis shifted from a situational to a personal exploration. A considerable amount of historical data emerged and frequently provided abreactive and cathartic responses. The recall of forgotten memories was often accompanied by violent emotional reactions.

The group of impotent men displayed a high degree of empathy for one another, and frequently expressed feelings of hostility and resentment towards the therapist. These feelings were accepted by him and clarified for the patients; they were followed by discussions of the effects of frustration.

At the end of each session, the therapist provided a summary of the proceedings. He attempted not only to recapitulate the remarks of the subjects and to reflect back to the group the emotional significance of their statements, but also to suggest possible connections between their symptoms and their feelings. Interpretive remarks dealt mainly with possible motives behind the façade of manifest behavior. Premature interpretations were vigilantly avoided. Obvious rationalizations, as well as statements of overprotestation, were challenged by the therapist only when he sensed a readiness on the part of the group.

Both the interpretive groups and the desensitization sessions were usually conducted three times a week. The desensitization groups were disbanded when all the patients were able to tolerate the most severe anxiety producing stimulus in the hierarchy without undue disturbance.

Members of the interpretive groups were given the same number of sessions as the corresponding desensitization groups. Since very few patients recovered from their phobias by means of the interpretive procedures, the ones whose phobic symptoms persisted were provided with an opportunity of undergoing group desensitization. (There were too few desensitization failures to satisfy the minimum numerical requirements for comparable interpretive groups.) Although the main response was an immediate willingness to introduce the "different group technique," the group of impotent men decided to continue employing interpretive procedures a while longer. Group desensitization was then administered to those patients who were not rendered symptom-free by the interpretive methods.

RESULTS

Assessment of Recovery

One month after therapy had terminated, the acrophobic and claustrophobic patients who claimed to have recovered from their phobias were required to undergo additional stress tolerance tests.

The acrophobic subjects were required to climb to the third landing of a fire escape (a height of about 50 feet). From the third story, they were required to go by elevator with the experimenter to the roof garden, eight stories above street level, and then to count the number of passing cars for 2 minutes.

The claustrophobic subjects were required to remain in the cubicle with the French windows shut and the movable screen a few inches away. Those who were able to endure the situation with no apparent distress for 5 minutes were regarded as recoveries, provided that they were also able to present satisfactory evidence that they were no longer handicapped in their life situations. The tests were conducted individually in the presence of a witness.

With two exceptions, all the patients who stated that they had recovered from their phobias were able to face the tolerance tests with outward tranquility, although some of the acrophobic patients later admitted that they had felt "a trifle anxious" when looking down from the edge of the roof garden.

Neither the impotent men nor the members of the mixed phobic group were objectively tested.

The most rigorous criteria were used in assessing therapeutic results. For instance, only those patients who displayed an unambiguous posttherapeutic freedom from their respective phobic disorders were classified as recoveries. These criteria were, of course, essentially symptomatic. If a claustrophobic patient for instance, was still unable after therapy to visit the cinema for fear of suffocation, his treatment was considered a failure, regardless of any ex parte testimony to the contrary. Merely to enable a patient to "accept his neurosis" or to achieve a so-called "personality reintegration" without symptomatic relief was considered not good enough. Recovery from a phobic condition implies total neutrality or indifference to the formerly anxiety generating stimulus constellation. The present study made no provision for moderate or slight improvements. The latter were all classified as failures.

TABLE 1
NUMBER OF PATIENTS ASSIGNED TO EACH CONDITION AND THE THERAPEUTIC OUTCOME

Patients	Treated by desensitization	Recovered	Treated by interpretation	Recovered	Treated by interpretation and relaxation	Recovered
Acrophobics	5	4	3	0	3	1
Claustrophobics	7	4	3	0	5	1
Impotence	2	2	3	0	—	—
Mixed group	4	3	—	—	—	—
Total	18	13	9	0	8	2

Statistical Analysis of the Results

Results are summarized in Table 1. As shown, there were 13 recoveries and 5 failures for desensitization, 2 recoveries and 15 failures for other forms of treatment. The resulting chi square is 10.69, which is highly significant ($p < .01$).

Additional statistical comparisons were computed for the matched pairs of acrophobic and claustrophobic patients who received group desensitization or group interpretation, respectively. There was a total of five matched pairs in these groups. Both members remained unimproved in one pair; in four, the desensitization patients recovered but the interpretation cases failed. There were no pairs in which both recovered or in which only interpretive methods succeeded. By applying the null hypothesis that the two methods are equally effective, the probability of obtaining this result is .0625.

In the case of the impotent men, no matching was carried out. Fisher's (1946, p. 97) exact test, which gave a probability of .1, was employed for testing significance. When the two probabilities, .0625 and .1, were combined (Fisher's 1946, p. 99), the resulting level of significance was .03, favoring desensitization.

It is interesting to note that when comparisons were made between the matched pairs of acrophobic and claustrophobic patients who received group desensitization as opposed to group interpretation-plus-relaxation, the level of significance in favor of desensitization

was only 12.5%. There were three matched pairs in which desensitization proved successful and interpretation-plus-relaxation failed. There were no cases where group interpretation plus-relaxation succeeded while group desensitization failed. Both methods failed with three matched pairs and both methods succeeded with one matched pair.

Since a significance level of 12.5% falls outside the conventional limits, the obvious conclusion is that there is no evidence of differences between the desensitization and interpretation-plus-relaxation conditions. Of course, seven matched pairs provides one with little leverage and the statistical analysis of such a small number cannot be conclusive. It is worth noting, however, that of the six individuals who were initially unsuccessfully treated by interpretation-plus-relaxation, four later recovered from their phobias after a mean of 9.8 group desensitization sessions.

Of the total of 15 patients who had derived no apparent benefit from the interpretive procedures, 10 recovered from their phobias after a mean of 10.1 group desensitization sessions, as compared with the mean of 20.4 sessions which were necessary for effective group desensitization when only this procedure was employed.

Follow-Up Studies

Follow-up studies were conducted by means of the following questionnaire:

1. Has your original phobic disorder returned?
2. If you have had a relapse, is it slight, moderate, or severe?
3. Since receiving treatment have you developed any new symptoms? (If so, please elaborate.)
4. Please underline all the following complaints which apply to you:
Tension Depression Anxiety Palpitations Dizziness Insomnia Nightmares Headaches Tremors Sexual problems Fatigue Stomach trouble Other symptoms (specify)
5. Please indicate whether any of the above complaints commenced *after* your participation in the therapeutic groups.
6. Are you still handicapped in any area of your daily living? (Specify.)
7. Have you consulted another therapist?

The duration of after-study history varied from group to group, and ranged from 15 months to 1.5 months with a mean of 9.05 months. All those subjects whose follow-up reports revealed even slight phobic recurrences were considered to have relapsed. Particular attention was devoted to the question of possible symptom substitution, but no evidence of this phenomenon was encountered.

When the follow-up evaluations were taken into account, 10 of the 13 patients who had recovered by means of group desensitization still maintained their freedom from phobic symptoms. Thus, 3 patients were regarded as having relapsed.

Of the 2 patients who had recovered after undergoing group interpretation-plus-relaxation, 1 maintained his recovery.

Eight of the 10 patients who recovered after undergoing postinterpretive group desensitization maintained recovery.

Summary of Findings

Group desensitization was applied to 18 patients of whom 13 initially recovered and 3 subsequently relapsed.

Group interpretation was applied to 9 patients. There were no recoveries in this group.

Group interpretation-plus-relaxation was applied to 8 patients of whom 2 recovered and 1 subsequently relapsed.

The 15 patients who had not benefited from the interpretive procedures were then treated by group desensitization. There were 10 recoveries of whom 2 subsequently relapsed.

DISCUSSION

Wolpe (1958) has expressed the basis of his "reciprocal inhibition" therapy as follows:

> If a response incompatible with anxiety can be made to occur in the presence of anxiety-evoking stimuli so that it is accompanied by a complete or partial suppression of the anxiety-responses, the bond between these stimuli and the anxiety-responses will be weakened.

His method of systematic desensitization based on relaxation incorporates Jacobson's (1938) finding that muscular relaxation inhibits anxiety and that their concurrent expression is physiologically impossible.

The deliberate use of the parasympathetic accompaniments of skeletal muscular relaxation to inhibit neurotic anxieties reciprocally may be termed "specific reciprocal inhibition." There is, however, a broad range of stimuli which have *nonspecific* properties for inhibiting neurotic responses reciprocally. The more

usual clinical medium of verbal interchange, for instance, may in itself bring about the incidental or nonspecific reciprocal inhibition of neurotic responses. In other words, it is postulated that *interview situations* can sometimes evoke autonomic responses similar to those of deep muscle relaxation.

The fact that far fewer sessions were required to desensitize those subjects in the present sample who had previously received interpretive therapy may be explicable by the notion that interpretive group situations evoked appropriate emotional responses in most of the subjects to inhibit some of their anxieties. In other words, it is probable that some of the anxiety responses evoked by the group discussions underwent a measure of nonspecific reciprocal inhibition. Furthermore, those patients who received postinterpretive group desensitization had the advantage of having established a therapeutic relationship with the experimenter. It is postulated that "the therapeutic atmosphere of empathy and acceptance may in itself reciprocally inhibit neurotic anxieties" (Lazarus, 1959).

It should be mentioned that the interpretive groups apparently enabled many of the patients to achieve a constructive modification of their self-evaluation, often clarified their evaluation of others, and enhanced their potentialities of interpersonal integration. These gains, however, appeared to have little bearing on their phobic symptoms, which usually persisted until desensitization procedures were administered.

The comparatively high relapse rate in the present series is probably related to the fact that the treatment was rather narrowly confined to a single range of stimuli which could in some cases have been a small part of a broad constellation, other elements of which may have afforded additional and possibly more useful bases for desensitization. In a proper clinical setting, the group desensitization procedures would have been preceded by individual history taking and the compilation of detailed clinical information for use either individually or in the group situations. Consequently, the conditions for the application of desensitization therapy were far less than optimum, a point which suggests that the experimental outcomes are only minimally indicative of the utility of this therapeutic approach to phobic symptoms.

The concept of "experimenter bias" is a relevant consideration in any study of this kind. It is difficult to determine the extent to which the present results were influenced by the therapist's theoretical affiliations. In terms of subjective interest, however, it should be noted that the experimenter's preferences were decidedly in favor of the interpretive methods. Fortunately, the ennui which is generated while applying desensitization procedures is adequately offset by the gratifying results.

If another therapist had treated the interpretive groups, a significant difference in the results might merely have reflected the superiority of the individual therapist rather than the methods employed. The treatment of phobias by interpretive methods, however, is well known to be difficult. Curran and Partridge (1955), for instance, state that "phobic symptoms are notoriously resistant to treatment, and their complete removal is rarely achieved." Similar views are expressed by Maslow and Mittelmann (1951), Henderson and Gillespie (1955), and Mayer-Gross, Slater, and Roth (1955). By contrast, phobias respond to desensitization exceedingly well (Eysenck, 1960; Lazarus & Rachman, 1957; Wolpe, 1958). It is contended, therefore, that the superior results achieved by group desensitization are not a function of the therapist's disproportionate skills (or unconscious prejudices) but a reflection of the intrinsic value of desensitization per se in the treatment of phobic disorders.

The point may legitimately be raised as to whether desensitization achieves any result other than the elimination of the phobic symptom. Comments on the general repercussions of desensitization are not possible in the context of the present study. No attempt was made to study changes in personality or general adaptation. Many patients, however, made remarks which suggested that the elimination of a phobic symptom is not an isolated process, but has many diverse and positive implications. As Eysenck (1959) states:

> The disappearance of the very annoying symptom promotes peace in the home, allays anxiety, and leads to an all-round improvement in character and behavior.

The extent to which desensitization is a method of *general* applicability (i.e., whether this method would benefit any neurotic patients other than those suffering from

phobic disorders) is also worthy of mention. The value of desensitization is limited to those conditions wherein appropriate hierarchies can be constructed and where specific rather than pervasive anxiety is present. In other words, it is only where reasonably well-defined stimulus configurations can be identified that desensitization techniques should be applied. For example, patients whose interpersonal relationships are clouded by specific fears of rejection, hypersensitivity to criticism, clear-cut areas of self-consciousness, or similar specific anxiety evoking stimuli often derive benefit from desensitization procedures. By contrast, desensitization cannot readily be applied in such cases as character neuroses, hysterical disorders, and chronic inadequacy. A further prerequisite for the effective application of desensitization is the ability to conjure up reasonably vivid visual images which elicit emotional reactions comparable to the feelings evoked in the real situation.

While dealing with the limitations of desensitization procedures, one should not lose sight of the fact that systematic desensitization appears to be a most valuable technique in the alleviation of phobic disorders. The fact that this method can be effectively administered in *groups* suggests greater availability with little loss in economy or effectiveness for phobic sufferers.

SUMMARY

Wolpe's (1958) technique of systematic desensitization based on relaxation was adapted to the treatment of phobic disorders in groups. Of the 18 subjects who were treated by direct group desensitization, 13 recovered in a mean of 20.4 sessions. Follow-up inquiries after an average of 9.05 months revealed that 3 of the subjects had relapsed. With a more traditional form of interpretive group psychotherapy applied to 17 subjects, after a mean of 22 therapeutic meetings, it was found that only 2 patients were symptom-free. Both these patients had attended groups in which relaxation was employed as an adjunct to the interpretive procedures. The 15 subjects who were not symptom-free after interpretive group therapy were then treated by group desensitization. After a mean of 10.1 sessions, 10 of them recovered. The very much shorter time required to effect a recovery by desensitization in those patients who had previously received interpretive therapy suggests that the therapeutic relationship and additional nonspecific factors may have facilitated the reciprocal inhibition of neurotic anxieties motivating the phobic symptoms. There is some basis for the idea that therapists of every persuasion could helpfully employ systematic desensitization as an adjunct to their traditional techniques in the management of phobic disorders.

REFERENCES

CORSINI, R. J. *Methods of group psychotherapy.* New York: McGraw-Hill, 1957.

CURRAN, D., & PARTRIDGE, M. *Psychological medicine.* Edinburgh & London: Livingstone, 1955.

EYSENCK, H. J. Learning theory and behavior therapy. *J. ment Sci.*, 1959, **105**, 61–75.

EYSENCK, H. J. (Ed.) *Behaviour therapy and the neuroses.* Oxford: Pergamon, 1960.

FISHER, R. A. *Statistical methods for research workers.* Edinburgh: Oliver & Boyd, 1946.

HENDERSON, D., & GILLESPIE, R. D. *A text-book of psychiatry.* London: Oxford Univer. Press, 1955.

JACOBSON, E. *Progressive relaxation.* Chicago: Univer. Chicago Press, 1938.

LAZARUS, A. A. The elimination of children's phobias by deconditioning. *S. Afr. med. Proc.*, 1959, **5**, 261–265.

LAZARUS, A. A., & RACHMAN, S. The use of systematic desensitization in psychotherapy. *S. Afr. med. J.*, 1957, **31**, 934–937.

MASLOW, A. H., & MITTELMANN, B. *Principles of abnormal psychology.* New York: Harper, 1951.

MAYER-GROSS, W., SLATER, E., & ROTH, M. *Clinical psychiatry.* London: Cassel, 1955.

WOLBERG, L. R. *The technique of psychotherapy.* New York: Grune & Stratton, 1954.

WOLPE, J. *Psychotherapy by reciprocal inhibition.* Stanford: Stanford Univer. Press, 1958.

(Received September 22, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 63, No. 3, 511–516

SOME PARAMETERS OF THE CONDITIONING OF VERBAL AFFECT RESPONSES IN SCHIZOPHRENIC SUBJECTS[1]

KURT SALZINGER AND STEPHANIE PISONI

New York State Department of Mental Hygiene and Columbia University

THE application of operant conditioning techniques to verbal behavior has proved fruitful in a variety of situations (see recent reviews by Krasner, 1958; Salzinger, 1959b). The specific task of the experimental analysis of the interview has been outlined by Salzinger (1959a), and this approach has met with experimental success.

It has been shown that the output of verbal affect responses of schizophrenic patients can be manipulated in an interview situation by the use of discriminative stimuli such as questions (Salzinger, 1956; described in Zubin, 1958) and by the use of reinforcement such as verbal expressions of agreement (Salzinger, 1956; Salzinger & Pisoni, 1958; described in Zubin, 1958). Verbal reinforcement was also shown to be effective in increasing the frequency of occurrence of verbal affect in normal subjects (Salzinger & Pisoni, 1960), and a comparison of a schizophrenic and a normal group revealed a difference only in their extinction rates.

In the following studies a number of additional parameters of the conditioning of verbal affect in an interview situation were investigated.

EXPERIMENT I

The following study[2] was undertaken to compare the effectiveness of verbal reinforcement when given at the beginning of the interview with its effectiveness when given in the middle.

[1] This study was supported by Research Grant M586 (C3) and M1541 from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service.

The authors are indebted to J. Zubin and S. Sutton for their help and interest in this study. The authors are also grateful to N. Beckenstein, Director of the Brooklyn State Hospital, to A. Glenn and L. Olinger, and to M. Portnoy for their cooperation in helping to make available the sample used for this study, and to R. S. Feldman and A. Reisel for their help in analysis of the data.

[2] This paper was presented in part at the American Psychological Association, New York City, 1957.

Procedure

Subjects for this experiment were 10 female and 4 male hospitalized schizophrenics, ranging in age from 21 to 48 years with a median of 34. Subjects were selected from admissions to a state hospital and were accepted as schizophrenic on the basis of the diagnosis given at the hospital distribution center. Eight patients had been previously hospitalized. None had received insulin, electroshock, or tranquilizing drugs for at least one week before the interview.

Each subject was interviewed one week after arrival in the hospital. All interviews were recorded with the recording equipment in full sight of both the subject and the experimenter.

The interview was presented to the subject as a routine mental hospital procedure. The experimenter brought the subject into the experimental room and explained that the interview was being conducted to help him.

The experimenter questioned the subject about his name, age, marital status, children, and siblings. The subjects had little difficulty answering these questions. Any response given was reinforced by the experimenter repeating the answer, writing it down, and saying "mm-hmm," "uhha," or "I see." This procedure was adopted in an effort to obtain factual information upon which the remaining interview questions could be based and to establish the experimenter as a source of reinforcement, thus encouraging the subject to speak in the presence of the experimenter. Then the experimental part of the interview was initiated with the question, "Would you tell me why you are here in this hospital?"

All subjects were interviewed for a period of 30 minutes. During the first 10-minute period (conditioning), the experimenter questioned the subject and reinforced each affect response (as defined below). By reinforcement is meant verbal agreement by the experimenter through the use of such words as "mm-hmm," "uhha," "I see," "yeah," "yes," etc. immediately following the affect statement.

This was followed by 20 minutes of extinction (divided into two 10-minute periods in the analysis presented below). The experimenter continued asking questions but withheld all reinforcement.

Definition of Response

The response class of affect was defined as any statement by the subject describing or evaluating his own state in other than intellectual or physiological terms. In other words, included in the response class were only those affect statements in which the subject referred the affect to himself by using the pronouns I or We, as indicated in the following examples: "I am

satisfied," I'm happy," "We enjoyed it," "I like him" (Salzinger & Pisoni, 1958).

A count per minute was taken of affect statements and general self-referred statements (all statements beginning with I or We). The latter served as a measure of each subject's general output of self-referred statements.

Interviewer Questions

After the question, "Why are you here?" the experimenter asked additional questions only when the subject ceased talking for at least 2 seconds. The following topics were discussed during the interview: reasons for being in the hospital, causes for illness, patient's relationships to his parents, siblings, fellow employees, employers, fellow students and teachers, wife or husband, children and friends, patient's activities during free time, and plans for the future. The experimenter made an attempt to balance these topics over the different conditions of the interview by bringing up the same topics in each 10-minute period. As long as the topics were approximately balanced over the three 10-minute periods, however, the experimenter took his cues as to topic from the content of the subject's statements. Thus both the number of topics and the order in which they were discussed varied from interview to interview. Questions dealing directly with affect—e.g., "How did you feel about that?" or "Were you happy?"—were not used.

The results obtained from this group (Group B) were compared with those obtained from two previously reported groups (Salzinger & Pisoni, 1958). For one group (Group A) the 30-minute interview was divided in the following way: During the first 10 minutes (operant level), the base rate of affect statements was determined—i.e., the experimenter did not reinforce any statement; during the second 10 minutes (conditioning), the experimenter reinforced all affect statements; during the third 10 minutes (extinction), the experimenter again withheld reinforcement. For the second group (Group C), *no* reinforcement was given for the entire 30 minutes of the interview.

Results

Conditioning effect for total groups. Friedman's nonparametric analysis of variance (Siegel, 1956) was applied separately to each group in order to determine whether reinforcement produced an increase in affect statements in the period where it was delivered. Inspection of these total group analyses in Table 1 shows that the largest number of affect responses occurred during conditioning when the reinforcement was given in the middle of the interview (Group A).[3] There is a nonsignificant concentration of affect statements in the first 10 minutes when the reinforcement was given in the beginning of the interview (Group B). The control group (Group C) shows a nonsignificant clustering of affect statements in the second 10-minute period of the interview.

TABLE 1

SUM OF RANKS OF AFFECT STATEMENTS FOR EACH OF THE TEN-MINUTE PERIODS FOR THE THREE GROUPS OF PATIENTS

Group	Successive ten-minute periods			χ^2_r	p	N
	First	Second	Third			
A	41.5	50.5	28.0	12.82	.002	20
B	31.5	25.5	27.0	1.39	.50	14
C	30.0	37.0	29.0	2.37	.30	16

Conditioning effect as a function of a minimum number of reinforcements. A comparison of the number of reinforcements given to Groups A and B reveals that members of the former received a slightly larger number of reinforcements than the latter (although $p = .50$). Furthermore, examination of Group C (which received no reinforcement) shows a distribution ($p = .30$) in a direction opposite to that expected from the experimental paradigm of Group B and in the same direction as that expected in Group A. Although neither of these two differences is statistically significant, the combination of both trends would tend to obscure the effect of the reinforcement in Group B for two reasons: Conditioning may not take place if the number of reinforcements is too small, and the expected larger number of affect statements in the first 10 minutes may not be apparent because of the "natural" trend (see Group C) for the second 10 minutes to exceed the first 10 minutes.

For these reasons, and also because a previous study (Salzinger & Pisoni, 1958) had already shown that the strength of a response varies as a function of the number of reinforcements applied to it, it was decided to search for the minimum number of reinforcements necessary for conditioning to take place.

[3] For this paper, all interviews were carefully typed and later checked by the investigators to insure the accuracy of the frequency counts. In the course of this process, small changes were made in the coding of the interviews previously reported (Salzinger & Pisoni, 1958), thus producing discrepancies between the sums of ranks reported in that paper and in this study. It is to be noted, however, that the new counts showed no difference in the substantive results; all the conclusions as reported then are still supported by the corrected data.

This was accomplished by successively removing from Group B those individuals who received the lowest number of reinforcements until the group showed the conditioning effect. When those subjects receiving the lowest number of reinforcements (two) were removed, the sums of ranks for each of the succeeding 10-minute periods (conditioning period and first and second extinction periods) were 28.5, 22.5, and 21.0, respectively ($p = .27$). When the individuals having the next highest number of reinforcements (four) were removed, the sums of ranks changed to 25.0, 18.5, and 16.5 ($p = .15$). Finally, when subjects receiving the next highest number of reinforcements (six) were removed, the sums of ranks were 22.0, 15.5, and 10.5 ($p = .02$). In other words, subjects receiving six reinforcements or less did not give evidence of conditioning, whereas subjects receiving a larger number did.

A reasonable question arising from this analysis is whether the selection of individuals with a larger number of reinforcements is not simply a matter of biased sampling. Consequently, the equivalent procedure was applied to Group C, which received no reinforcement. Removing subjects with less than seven affect statements in the first 10 minutes produced the following sums of ranks: 11.5, 15.0, and 9.5 ($p = .27$). No significant difference, then, was found between the periods of the interview, and the trend was quite different from the one found in Group B after removing subjects who were given less than seven reinforcements.

Comparison of conditioning and extinction effects in Groups A and B. A comparison of the number of affect statements emitted during the conditioning periods of Groups A and B showed no significant difference ($p = .72$). A comparison of the number of affect statements emitted during the last 10 minutes of Group A (extinction) with the second 10 minutes of Group B (first extinction period) yielded no significant difference ($p = .50$), nor was there a significant difference ($p = .22$) between the extinction period of Group A and the last 10 minutes of Group B (the second extinction period).

In view of the fact that conditioning was demonstrated in Group B for only those individuals who received more than six reinforcements, Group A was subjected to a similar analysis by removing all subjects who had received six reinforcements or less during conditioning. The resulting sums of ranks for the three conditions were operant level, 26.0; conditioning, 38.0; and extinction, 20.0 ($p = .002$). Inspection of the individuals who had received six or less reinforcements showed that most of these individuals did not condition. In other words, for Groups A and B, more than six reinforcements seemed necessary for conditioning.

Self-referred statements. Application of Friedman's analysis of variance to the change in the general class of self-referred statements revealed no statistically significant difference among the three 10-minute periods for Group C ($p = .44$). There were significant differences for Group B ($p = .03$) and for Group A ($p = .01$). The largest number of self-referred statements occurred during the first 10 minutes for both groups.

The Friedman analysis of variance was then applied to the general self-referred statements of those subjects in Groups A and B who had received more than six reinforcements. This analysis yielded no significant difference among the three 10-minute periods ($p = .12$ for Group A, $p = .45$ for Group B). In Group C, those individuals who emitted more than six affect responses in the first 10 minutes showed no significant difference in self-referred statements among the three 10-minute periods ($p = .35$).

Discussion

Although Group A emitted a larger number of affect statements than Group B during conditioning, the difference was not significant. Since Groups A and B showed significant conditioning effects, taking into account the number of reinforcements, it must be concluded that the effectiveness of reinforcement is the same whether the period of conditioning is placed at the beginning or in the middle of the interview. Different parts of the interview assume their importance not because of their temporal placement, but rather because of the presence or absence of reinforcement.

The procedure of selecting those subjects who received more than a minimum number of reinforcements to investigate the effectiveness of reinforcement was based on the assumption that a variable must be sufficient in magnitude

to be effective. Furthermore, the fact that the control group showed no "conditioning" when subjects are selected in an analogous manner suggests that a minimum number of reinforcements must be administered before a conditioning effect can be expected.

In these interview studies, however, the period of time during which reinforcement was available was held constant (10 minutes), and the number of reinforcements received was therefore dependent, in part, upon the subject's "normal" (operant level) output of verbal affect.

EXPERIMENT II

In a second experiment,[4] a given number of reinforcements was arbitrarily assigned to each subject prior to his entering the experimental situation. In this way the number of reinforcements administered was made independent of the subject's output. Furthermore, by allowing the duration of the conditioning period to vary, it was possible to study the rate of acquisition—i.e., the time between successive affect statements during conditioning—as another parameter in the conditioning of verbal affect.

Procedure

Subjects were 22 female and 10 male hospitalized schizophrenics, ranging in age from 19 to 51 with median age of 32.5 years. Eighteen subjects had been previously hospitalized. The selection of subjects was the same as in Experiment I.

The interview procedure was identical with that of Experiment I except that during the first 10 minutes (operant level), the base rate of verbal affect was determined. The experimenter asked questions but did not reinforce any statement made by the subject. During the second part of the interview (conditioning), the experimenter continued asking questions whenever necessary, but reinforced only a predetermined number of successive self-referred affect statements. Each subject was arbitrarily assigned a specific number of reinforcements: 4, 6, 8, 10, 15, or 20. The length of the conditioning period depended, therefore, upon the length of time it took the subject to receive his assigned number of reinforcements up to a maximum of 15 minutes.

Of the original 32 subjects, 17 received the assigned number of reinforcements in the allotted time, the conditioning periods ranging in length from 30 seconds to 14 minutes. The remaining 15 subjects did not receive their assigned number of reinforcements in the allotted time; therefore, their conditioning periods ran the full 15 minutes.[5]

Immediately following the conditioning period, all subjects were given 10 minutes of extinction during which the experimenter withheld all reinforcement.

Results

Conditioning effect as a function of minimum number of reinforcements. The analyses were restricted to those subjects who received the assigned number of reinforcements within the allotted time. Subjects who did not fall within this category were excluded because the number they did receive was determined not by the experimenter's assignment but by their operant levels. This was done in order to exclude operant level as a determining factor for number of reinforcements.

The mean number of affect statements per minute was found for each condition. The group was then divided into subjects who received 10 or more reinforcements and those who received 8 or less. A Friedman nonparametric analysis of variance, applied to the group receiving 10 or more reinforcements, revealed a significant difference among the three conditions of the interview ($p = .05$). The largest mean number of affect statements occurred during the conditioning period (sum of ranks = 21.5), the next largest during extinction (sum of ranks = 14.5), and the smallest during operant level (sum of ranks = 12.0).

When the Friedman test was applied to the three conditions of the interview for the group receiving eight or less reinforcements, no significant difference was found ($p = .40$). Examination of the trend revealed that the largest number of affect statements occurred during the extinction period (sum of ranks = 21.0), the next largest during conditioning (sum of ranks = 17.5), and the smallest during operant level (sum of ranks = 15.5).

When the same statistical procedures were applied to the general self-referred statements,

[4] This paper was presented in part at the Eastern Psychological Association, Philadelphia, 1958.

[5] The 2 subjects assigned to the 4-reinforcement group and the 6 subjects assigned to the 6-reinforcement group received their allotted number of reinforcements. For the 4 subjects assigned to the 8-reinforcement group, one did not get the allotted number; of the 6 assigned 10 reinforcements, 3 did not; of the 7 assigned 15 reinforcements, 5 did not; of the 10 assigned 20 reinforcements, 6 did not.

Comparisons of the subjects who received their assigned number of reinforcements with those who did not revealed no statistically significant differences ($p > .05$) in terms of sex, age, educational level, or admission status.

no statistically significant differences were found either for subjects who received 10 or more reinforcements ($p = .65$) or for subjects who received 8 or less ($p = .38$).

Number of reinforcements and resistance to extinction. In order to investigate the relationship between number of reinforcements and resistance to extinction, a Spearman rank-order correlation coefficient (Siegel, 1956) was computed between number of reinforcements and the number of affect statements in the extinction period. For the total group, the correlation was not significant ($r_s = +.29$, $p > .05$). For subjects with 10 or more reinforcements, the correlation between number of reinforcements and number of affect statements in the extinction period was $-.04$ ($p > .05$); for subjects who received 8 or less, the correlation was $+.68$ ($p < .05$).

Rank-order correlations were also computed for the relation between the rate of acquisition and resistance to extinction. The ratio of number of affect statements to the time taken to emit them was correlated with the number of affect statements during extinction. For the total group, the correlation was $+.28$ ($p > .05$); for subjects receiving 10 reinforcements or more, the correlation was $+.64$ ($p < .05$); for subjects receiving 8 or less, the correlation was $+.22$ ($p > .05$).

To investigate further the effect of rate of acquisition upon resistance to extinction, 11 subjects from this sample were matched with 11 subjects selected from a previous study (Salzinger & Pisoni, 1958). The interview procedures were identical except that the length of the conditioning period in the earlier study was held constant (10 minutes) and the number of reinforcements administered varied as a function of the subject's verbal behavior. These 11 pairs were matched in terms of diagnosis (schizophrenic), sex, operant level (no difference greater than four responses in 10 minutes), and number of reinforcements received (no difference greater than two). They varied only in terms of their rate of acquisition, i.e., time taken to receive the same number of reinforcements.

Nine of the 11 subjects who had the faster rate of acquisition also showed a larger number of extinction responses. A paired replicates test (Siegel, 1956) of these differences was significant ($p = .03$). The median difference in number of extinction responses was seven.

Discussion

In this experiment, where the number of reinforcements administered was independent of the subject's behavior, it is clear that a minimum number of reinforcements is necessary for conditioning to take place. Subjects who are given less than this minimum number generally show no evidence of conditioning. This result corroborates the findings in Experiment I.

Allowing the duration of the conditioning period to vary permitted an evaluation of the parameter of rate of acquisition. It appears that number of reinforcements, within the range used in this experiment, in and of itself, does not relate to resistance to extinction. A more important factor is the time taken to receive these reinforcements, i.e., the rate of acquisition. When individuals were matched in terms of their operant levels and the number of reinforcements received, but allowed to differ in the rate of acquisition, it was shown that the faster the acquisition, the greater is the resistance to extinction. Very likely, as the number of reinforcements given increases, the rate of acquisition becomes a less important factor in determining resistance to extinction.

This finding, however, has importance for the large number of brief interviews utilized by many psychologists. The interviewer who uses verbal reinforcement influences the interviewee's behavior, but his delivery of reinforcement is influenced in turn by the patient's behavior. This implies that the amount of influence (bias) that a given interviewer exerts will vary differentially with the rate of acquisition of each particular subject. Before an investigator can say that two groups of subjects, whether patients or normals, differ from each other in terms of the information, opinion, or feeling he has elicited, he must be able to show not only that his reinforcement procedure is identical in both groups, but also that their reaction to reinforcement (rate of acquisition) is the same.

SUMMARY

The present studies were designed to investigate some parameters in the conditioning of verbal affect in an interview situation with schizophrenic subjects. In the first experiment it was shown that the process of conditioning verbal affect is independent of the

temporal placement of the conditioning period—at the beginning or in the middle of a 30-minute interview—and that the administration of a minimum number of verbal reinforcements is necessary for the conditioning effect to take place.

In a second experiment, where the number of reinforcements administered was independent of the subject's behavior and determined by the experimenter's arbitrary assignment, it was shown that conditioning of verbal affect statements occurs for those individuals who receive 10 reinforcements or more, but not for those who receive less. Furthermore, rate of acquisition was related to resistance to extinction; those subjects who had the faster rates of acquisition showed the greater resistance to extinction.

REFERENCES

KRASNER, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.*, 1958, **55,** 148–170.

SALZINGER, K. The experimental approach to the interview. In J. Zubin (Ed.), *Experimental abnormal psychology.* New York: Columbia Univer. Library, 1959. Pp. 1–46. (Mimeo) (a)

SALZINGER, K. Experimental manipulation of verbal behavior: A review. *J. gen. Psychol.*, 1959, **61,** 65–94.(b)

SALZINGER, K., & PISONI, STEPHANIE. Reinforcement of affect responses of schizophrenics during the clinical interview. *J. abnorm. soc. Psychol.*, 1958, **57,** 84–90.

SALZINGER, K., & PISONI, STEPHANIE. Reinforcement of verbal affect responses of normal subjects during the interview. *J. abnorm. soc. Psychol.*, 1960, **60,** 127–130.

SALZINGER, SUZANNE. Rate of affect responses in schizophrenics as a function of three types of interviewer verbal behavior. Paper read at the Eastern Psychological Association, Atlantic City, 1956.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

ZUBIN, J. A biometric model for psychopathology. In R. A. Patton (Ed.), *Current trends in the description and analysis of behavior.* Pittsburgh: Univer. Pittsburgh Press, 1958. Pp. 22–47.

(Received October 25, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 63, No. 3, 517-523

EYELID CONDITIONING AND THREE PERSONALITY PARAMETERS[1]

J. G. FIELD[2] AND J. C. BRENGELMANN[3]
Institute of Psychiatry, Maudsley Hospital, University of London

ACCEPTING the Hullian system, the Iowa group has argued that high drive subjects will, under conditions where a single habit tendency is evoked, have a higher performance level than low drive subjects. One measure of performance was derived from conditioning the eyelid response.

To measure drive, a questionnaire, the Taylor (1953) Manifest Anxiety (*MA*) scale was constructed. No attempt was made to make the *MA* unidimensional and there is strong evidence that it measures neuroticism (Eriksen & Davids, 1955; Kerrick, 1955) and, to some extent and negatively, extraversion (Franks, 1956b).

Five experiments comparing extreme scorers on the *MA* are construed as supporting the Iowa hypothesis (Spence & Farber, 1953; Spence, Farber, & Taylor, 1954; Spence & Taylor, 1951, 1953; Taylor, 1951). When the entire range of *MA* scores is used, the correlation with conditioning is .25 (Taylor, 1956).

Eysenck (1955) and Franks (1956a) questioned the construction placed on these experiments. Also basing themselves on Hull's system, they suggested that the Iowa findings could equally well be accounted for by a different explanation. Individual differences in the performance of conditioned responses may be due not to differences in drive, but to differences in the growth and decay of habit strength. This change in emphasis is supported by an experiment by Franks (1957a) in which no differences in eyelid conditioning were found between a satiated (low drive) student group and one deprived of food and water (high drive).

In an effort to unify diverse experimental findings in the psychology of personality, Eysenck (1957) has attempted to show that one major source of individual differences is attributable to differences in the excitation-inhibition balance of the central nervous system. Eysenck postulated that those individuals in whom excitatory potential is generated slowly and is weak and in whom reactive inhibition is developed quickly, is strong, and dissipates slowly, tend to exhibit extraverted behavior patterns; conversely, those individuals in whom excitatory potential is generated quickly and is strong and in whom reactive inhibition is developed slowly, is weak, and dissipates quickly, tend to exhibit introverted behavior patterns (p. 114).

A number of experiments have attempted to assess the relative importance of drive and extraversion-introversion in the contribution they make to the individual differences in eyelid conditioning. The measures of drive have been the *MA*, the Cycloid and Depression scales (Guilford, 1940), the Maudsley Medical Questionnaire (Eysenck, 1952), or the Maudsley Personality Inventory (MPI) Neuroticism scale (Eysenck, 1956). These scales are highly correlated with the *MA* and have been independently validated against psychiatric diagnoses. The measures of extraversion have been the Rhathymia scale (Guilford, 1940) and a development of it, the MPI Extraversion scale (Eysenck, 1956).

Two experiments support the hypothesis that it is the differences in the excitation-inhibition balance, as measured by extraversion scales, which make the more significant contribution to eyelid conditioning. The *MA* or equivalent scales fail to make a significant contribution, or do so to a significantly lesser degree (Franks, 1956a, 1957b). When subjects are unselected instead of in extreme groups, the correlation between extraversion (MPI) and conditioning is −.46 (acquisition) and −.34 (extinction trials) (Franks, 1957b).

Two other experiments provide consistent but not significant results favoring the Eysenck-Franks rather than the Iowa hypothesis (Das, 1957; Willett, 1960). In the in-

[1] The authors are indebted to Her Majesty's Prison Commission for permission to publish this article, but permission does not imply that it represents the Commissioner's views. The investigation was supported by the Medical Research Council and the Home Office.

[2] Now at Research Services Ltd., the London Press Exchange.

[3] Now at the State Department of Institutions and Agencies, State Colony, Woodbine, New Jersey.

vestigation reported below, a further attempt is made to assess the relative importance in eyelid conditioning of drive and excitation-inhibition balance as measured by the MPI Neuroticism and Extraversion scales. In addition, the contribution of "rigidity," a source of individual differences not previously considered in this context, is investigated.

Rigidity has been variously described as resistance to change, stereotypy, inflexibility, extremeness, dogmatism, intolerance, etc. Such concepts have been repeatedly criticized on empirical grounds and their generality questioned (Jenkins & Lykken, 1957). Brengelmann (1960c, 1960d) has followed up their suggestive importance with regard to learning. Apart from the concepts mentioned, achievement motivation ("drive") has been confirmed as one aspect of rigidity by the finding that scores on performance output (Brengelmann, 1960a) and aspiration level (Brengelmann, 1959a, 1960a) are correlated significantly with rigidity scales.

According to Brengelmann (1960a), the relationship between learning and rigidity depends on two effects of this personality syndrome: interference or inhibition, on the one hand, and drive or facilitation, on the other. The interference aspect has been repeatedly confirmed in the form of a negative relationship between performance and rigidity for perceptual-motor learning, reproduction learning, and immediate recall and recognition learning (Brengelmann, 1957, 1960a).

The drive effects require the assumption of a curvilinear relationship between rigidity and task difficulty (novelty, stress), similar to the Yerkes-Dodson law recently revived by Eysenck (1955) and Broadhurst (1959). According to this law, learning efficiency increases with task difficulty until a certain optimum level is reached and decreases thereafter. At the extreme levels of difficulty, drive, either absolutely or relatively, inhibits performance. This interpretation of rigidity has been confirmed in experiments using normal and abnormal subjects with regard to improvement with practice or output (Brengelmann, 1960a). For the high rigid subjects, improvement increases with difficulty of task up to a point of high difficulty and sharply decreases thereafter. For the low rigid subjects, the opposite effect is obtained. Medium levels of rigidity assume an intermediate position.

Many investigators equate a high rate of conditioned responses with good performance. Conditioning may also be considered as simple learning—learning *not* to respond to stimuli irrelevant, distracting, or interfering with respect to the task in hand. With respect to another more complex task, the same stimuli may be relevant. Situations favoring conditioning in the former situation will be detrimental in the latter. This view is supported by the generalization emanating from Brengelmann's work that whenever a personality variable correlates negatively with learning, its correlation with conditioning is found to be positive and vice versa. Accepting the widely practiced equation of eyelid conditioning with simple learning, rigidity is predicted as correlating positively and linearly with eyelid conditioning.

This prediction is consistent with the argument of Bindra, Paterson, and Strzelecki (1955) that the threshold to the blink is lowered by muscle tension, which, it has been suggested (Brengelmann, 1960b), underlies the excessive expressive movement found in the highly rigid.

The prediction may also be thought compatible with the Iowa hypothesis that higher drive favors the response with the greatest habit strength to the detriment of other responses. As applied to complex tasks, however, the argument loses consistency (Hill, 1957). Further, its predictions for a negative relationship between measures of drive and performance in learning are not significantly supported in Brengelmann's (1960a) experiments.

METHOD

The subjects of the experiment were 75 young men aged 25.81 $\pm$ *SD* 3.21 years. They were unselected samples of consecutive receptions at one of Her Majesty's prisons, where they were investigated at a research laboratory set up with the permission of the prison commissioners. These men had several previous convictions (6.45 $\pm$ 2.54), of which five-sixths were for property offences, and some previous experience of prison sentences (22.31 $\pm$ 1.59 months). About one-third of each sentence would have been remitted. Their intelligence, measured by a timed version of the Progressive Matrices (Raven, 1954), is normal in mean value and somewhat platykurtic in distribution.

Their occupations, if any, were mostly unskilled. Whether such subjects are more or less representative of the universe conventionally sampled by student volunteers or mental patients will not be discussed here.

The eyelid conditioning procedure of the Franks experiments was duplicated. The CS was a 1110 cps tone at about 65 db above threshold, established by the method of limits, of 800 msec. duration. After 350 msec. the UCS, a puff of air, lasting 500 msec., at a pressure of about 65 mm. of mercury at not less than 2 cms. from the right eye was administered by an electronic device. A partial reinforcement schedule was used, giving an overall reinforcement ratio of 60%. The intertrial interval varied between 20–30 seconds. To reduce the validity of the criticism that the individual differences in the performance of the blink are present *before* the acquisition trials, any subject blinking to any of six CS of nine pre-acquisition trials (3 CS, 3 UCS, 3 CS in that order) were excluded. This left 42 subjects. Franks (1956a) only excluded subjects if they blinked to either or both of the last two of these CS. The conditioning score was the total number of CRs over 18 test and 10 extinction trials between 150 and 625 msec. from the start of the CS. The maximum score possible was 28.

Three sets of four questionnaires each were used to assess the contribution of neuroticism, extraversion, and rigidity to individual differences in conditioning. The neuroticism and extraversion scales were administered in groups within 14 days of the conditioning. The rigidity scales were administered by post some months later. Retaining those who adequately completed the rigidity scales reduced the experimental group to 38 subjects. When the conditioning records were analyzed, it was found that five subjects had provided no measurable conditioned response. It is arguable that these subjects should be excluded because the absolute failure to condition may be due to lesions in the cns, misunderstanding, lack of cooperation, apparatus defect, or other extraneous factors. This would leave 33 subjects in the experimental group. The results are reported separately for the groups of 42, 38, and 33 subjects.

The four neuroticism questionnaires used were the MPI Neuroticism (N) scale, the Cattell second-order Anxiety (Anx) factor scale (Cattell, 1957), and the MMPI Psychasthenia (*Pt*) and *K* scales (Hathaway, 1946; Hathway & McKinley, 1943). High scorers on the *K* scale are expected to score low on neuroticism. These scales contained 24, 40, 48, and 30 items, respectively. The Anx, *Pt*, and *K* scales were selected from a factor analysis on the 42 subjects as having the highest loadings on a putative factor of neuroticism. The four questionnaires used to measure extraversion-introversion were the MPI Extraversion (E) scale, the second-order extraversion factor scale or Contact Personality Factor (CPF), Forms A and B summed (Cattell & King, 1955), the MMPI Hysteria (*Hy*) scale (Hathaway & McKinley, 1943), and the Activity Preference (AP) scale, designed to measure "anxiety" independently of neuroticism and shown to differentiate groups differing in conditioning performance (Lykken, 1957). Extraverts

TABLE 1

MEANS AND STANDARD DEVIATIONS OF PERSONALITY SCALES AND CONDITIONING FOR EXPERIMENTAL AND CONTROL GROUPS

	$M \pm SD$				
	Experimental group			Controls	
	(N = 42)	(N = 38)	(N = 33)	Group I (N = 25)	Group II (N = 80)
Neuroticism					
MPI N	25.12 ± 11.10	25.37 ± 11.37	25.33 ± 12.13	14.48 ± 9.40	20.99 ± 10.12
Anx	30.07 ± 10.56	30.65 ± 11.07	30.90 ± 11.52	—	26.75 ± 11.16[a]
Pt	13.60 ± 7.79	14.13 ± 8.33	14.61 ± 8.39	—	9.67 ± 6.22
K	10.83 ± 5.22	10.92 ± 5.38	10.94 ± 5.71	—	15.04 ± 4.23
Extraversion					
MPI E	27.43 ± 8.55	27.97 ± 8.78	27.79 ± 9.29	24.00 ± 7.25	29.30 ± 8.36
CPF	67.00 ± 10.69	68.00 ± 10.73	67.45 ± 11.37	—	65.94 ± 9.25
Hy	16.31 ± 4.97	16.82 ± 4.78	16.24 ± 4.54	—	16.81 ± 3.82
AP	19.69 ± 4.31	20.10 ± 4.15	20.24 ± 4.35	—	19.37 ± 4.22[b]
Rigidity					
CPIR	—	17.63 ± 5.27	17.27 ± 4.91	13.16 ± 4.91	—
NR	—	16.61 ± 5.43	16.15 ± 5.32	9.44 ± 6.26	—
IA	—	20.68 ± 4.54	20.67 ± 4.35	15.28 ± 6.80	—
Do	—	15.79 ± 5.88	16.00 ± 6.16	7.56 ± 7.10	—
Conditioning					
Acquisition	8.21 ± 5.16	8.13 ± 5.40	9.36 ± 4.69	—	6.40 ± 4.37
Extinction	3.69 ± 3.15	3.55 ± 3.23	4.09 ± 3.13	—	2.57 ± 2.65[c]
Total	11.62 ± 8.20	11.55 ± 8.38	13.30 ± 7.71	—	—

[a] From Cattell (1957), obtained on 759 normals.
[b] From Carpenter (1959), obtained on 73 United States servicemen.
[c] From Franks (1957a), obtained on 55 male students. Their Acquisition score was 7.18 ± 5.12.

TABLE 2

CORRELATIONS OF SCALES WITHIN EACH PERSONALITY PARAMETER AND WITH CONDITIONING

	Personality scales ($N = 38$)												Conditioning		
($N = 38$)	N	Anx	*Pt*	*K*[a]	E	CPF	*Hy*	AP[a]	CPIR	NR	IA	Do	($N = 42$)	($N = 38$)	($N = 80$)
N	—	56**	52**	59**									−04	−09	−11
Anx	57**	—	71**	60**									05	03	−01
Pt	56**	73**	—	62**									07	08	00
K[a]	59**	62**	70**	—									02	−04	−01
E					—	62**	21	40*					−08	−17	−17
CPF					64**	—	00	43**					−13	−15	−09
Hy					34*	−04	—	14					01	04	11
AP[a]					43*	42*	11	—					−06	06	12
CPIR									—	64**	19	29	—	15	33
NR									60**	—	19	29	—	09	27
IA									25	27	—	33*	—	06	09
Do									40*	34*	44*	—	—	43**	46**

[a] Sign reversed.

* Significant at beyond the 5% level.

** Significant at beyond the 1% level.

are expected to score low on this scale. These scales contained 24, 34, and 34, 60, and 32 items, respectively. The CPF, *Hy*, and AP scales were selected from the factor analysis mentioned as having the highest loadings on a putative factor of extraversion. None of the MMPI scales were T corrected.

Rigidity was measured by four questionnaires used in the Brengelmann studies (1959b, 1960c, 1960d). These were mainly derived from the California Personality Inventory Rigidity (CPIR) Scale (Gough), the Rigidity (NR) scale and Intolerance of Ambiguity (IA) scale (Nigniewitsky), and the Dogmatism (Do) scale (Rokeach). After item analysis by Brengelmann, these scales contained 15, 14, 17, and 16 items, respectively. The four scales have been found to correlate significantly and positively in a number of experiments (Brengelmann, 1959b; 1960c). In normal samples, rigidity does not, as a rule, correlate significantly with neuroticism and extraversion. In abnormal samples, rigidity and neuroticism are positively and significantly related, though to a slightly lesser extent than neuroticism and introversion.

RESULTS

Table 1 shows the values on the personality scales and on conditioning for the present experimental subjects. The data from two other groups (males) are included for comparison purposes. Group I ($N = 25$) resembles the experimental group in age (27.16 ± 4.26) and occupational level. Group II ($N = 80$) was composed of industrial apprentices in the age range 16–19 years (Willett, 1960). It is apparent from Table 1 that the experimental group was significantly more neurotic and rigid ($p < .01$) than Group I. The means in Group II are consistent with this finding.

Table 2 shows the relationships between the three personality parameters and conditioning for the experimental group. The neuroticism scales fail to account for any significant portion of the conditioning variance. Although the scales intercorrelate significantly and consistently, only some of the correlations with conditioning are in the direction expected on the basis of the Iowa hypothesis. The extraversion scales are not all significantly intercorrelated, nor entirely consistently. None significantly correlate with conditioning, and only some of them correlate in the predicted direction. Comparing the MPI N and E scales, however, E does contribute more to the conditioning variance than N in the direction predicted on the basis of the Eysenck-Franks hypothesis. All rigidity scales intercorrelate positively, but only partly significantly. Similarly, the correlations between the rigidity scales and conditioning are all consistent with the prediction, and dogmatism provides a highly significant correlation with conditioning, occupying about a fifth of the conditioning variance. For the acquisition trials alone, this correlation is .44 ($N = 38$).

As only one of the four rigidity scales

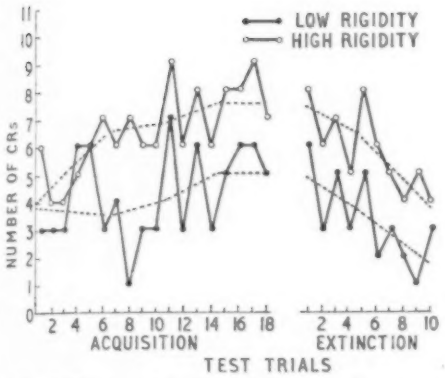

FIG. 1. Number of conditioned responses for groups high and low on four rigidity scales.

correlates significantly with conditioning, a multiple correlation was run between the four rigidity scales and conditioning. The multiple correlation was .46, significant beyond the 5% level (or .55, $N = 33$, $p < .05$).

Figure 1 presents the conditioning curves for acquisition and extinction for two subgroups high and low on rigidity, selected from the experimental group of 33 subjects. The high rigidity group comprises the 11 subjects who obtained the highest total scores summed over the four rigidity questionnaires. The low rigidity group comprises the 11 subjects who obtained the lowest total scores summed over the four questionnaires. The broken lines in Figure 1 represent the mean number of CRs, taken four at a time for each subgroup.

DISCUSSION

Of the three personality parameters, neuroticism, extraversion, and rigidity, only rigidity contributed a significant amount of variance to individual differences in eyelid conditioning. It should be remembered that the way in which the experiment was carried out was biased against these findings. The rigidity questionnaires were administered without supervision, when the subjects had been moved to other surroundings, and when considerable time since the rest of the testing had elapsed. Further, the rigidity scales were much shorter than the neuroticism and extraversion scales.

Should neither extraversion nor neuroticism measures reliably correlate with conditioning,

TABLE 3
COMPARISON OF SAMPLES AND EXPERIMENTS TESTING THE EYSENCK HYPOTHESIS

Experiment	Sample	Extraversion $M \pm SD$	Conditioning-Extraversion correlation
Franks (1956b)	20 Dysthymics	20.60 ± 6.29^a	−.15
Franks (1956b)	20 Hysterics	49.60 ± 11.22^a	−.40
Franks (1956b)	20 Normals	40.45 ± 11.63^a	−.05
Franks (1957b)	55 Students	26.18 ± 8.98^b	−.46*
Das (1957)	63 Students	25.58 ± 9.52^b	−.08
Willett (1960)	80 Adolescents	29.30 ± 8.36^b	−.08

[a] Rhathymia scale of Guilford (1940).
[b] Extraversion scale of Eysenck (1956).
* Significant at beyond the 1% level.

the original repeated findings with the *MA* can no longer be explained by the loading on extraversion noted by Franks (1956b). An alternative explanation is that the *MA* has a loading on rigidity. Maltzman, Fox, and Morrisett (1953) provide experimental support for this notion. Brengelmann (1960c) also found that rigidity correlated with the *MA* among neurotics. It may be that among normals the *MA* scores are too positively skewed for a correlation to appear (Wesley, 1953). On the other hand, although the correlation between extraversion and conditioning in the present investigation is not significantly different from zero, it is also not significantly different from the correlations found by Franks (1956a, 1957b). The samples and experiments testing the Eysenck hypothesis differ from each other in a number of respects, as is shown in Table 3. The somewhat unconventional setting of the present experiment can readily be supposed to influence the subject-experimenter interaction in a way that is untypical of university laboratories. There is, nevertheless, a consistent negative tendency in whatever relationship has been found between eyelid conditioning and extraversion questionnaires.

SUMMARY

The relative contributions to individual differences in eyelid conditioning of neuroticism and extraversion were assessed on the basis of the Spence-Taylor and the Franks-Eysenck hypotheses. In addition, a new parameter, rigidity, was predicted as positively related to conditioning on the basis of Brengelmann's

previous work. Personality was measured by 12 questionnaires. The Franks conditioning procedure was repeated, except that a more severe criterion of pre-acquisition blinking was used. Conditioning was measured by the number of blinks occurring to 28 tones, of which 18 were a part of a partial reinforcement schedule, and 10 were consecutive extinction trials. Seventy-five unselected normal male prison inmates were conditioned; 42 met the criterion, and 38 completed all the questionnaires. Only the rigidity measures provided a significant positive correlation with conditioning as predicted. The finding for extraversion and neuroticism was consistent with the Eysenck-Franks rather than the Spence-Taylor hypothesis, but not significantly so.

REFERENCES

BINDRA, D., PATERSON, A. L., & STRZELECKI, JOANNA. On the relation between anxiety and conditioning. *Canad. J. Psychol.*, 1955, **9**, 1–6.

BRENGELMANN, J. C. Extraversion, Neurotische Tendenz und Rigidität im Umkehrversuch (Prismenbrille). *Z. exp. angew. Psychol.*, 1957, **4**, 339–362.

BRENGELMANN, J. C. Abnormal and personality correlates of certainty. *J. ment. Sci.*, 1959, **105**, 142–162. (a)

BRENGELMANN, J. C. Differences in questionnaire response between English and German nationals. *Acta psychol., Amst.*, 1959, **16**, 339–355. (b)

BRENGELMANN, J. C. Bedingte Reaktionen, Lerntheorien und Psychiatrie. In H. W. Gruhle, R. Jung, W. Meyer-Gross, & M. Muller (Eds.), *Psychiatrie der Gegenwart*. Heidelberg: Springer, 1960. (a)

BRENGELMANN, J. C. Expressive movements and abnormal behaviour. In H. J. Eysenck (Ed.), *Handbook of abnormal psychology*. London: Pitman, 1960. (b)

BRENGELMANN, J. C. Extreme response set, drive level and abnormality in questionnaire rigidity. *J. ment. Sci.*, 1960, **106**, 171–186. (c)

BRENGELMANN, J. C. A note on questionnaire rigidity and extreme response set. *J. ment. Sci.*, 1960, **106**, 187–192. (d)

BROADHURST, P. L. The interaction of task difficulty and motivation: The Yerkes-Dodson law revived. *Acta psychol., Amst.*, 1959, **16**, 321–338.

CARPENTER, G. An investigation of the interrelationships between personality factors, conditions of difficulty and stress, and performance on a verbal learning task. Unpublished doctoral dissertation, University of London, 1959.

CATTELL, R. B. *The IPAT Anxiety Scale*. Champaign, Ill.: Institute for Personality and Ability Testing, 1957.

CATTELL, R. B., & KING, J. *The Contact Personality Factor Questionnaire*. Champaign, Ill.: Institute for Personality and Ability Testing, 1955.

DAS, J. P. An experimental study of the relation between hypnosis, conditioning and reactive inhibition. Unpublished doctoral dissertation, University of London, 1957.

ERIKSEN, C. W., & DAVIDS, A. The meaning and clinical validity of the Taylor anxiety scale and the hysteria-psychasthenia scales from the MMPI. *J. abnorm. soc. Psychol.*, 1955, **50**, 135–137.

EYSENCK, H. J. *The scientific study of personality*. London: Routledge & Kegan Paul, 1952.

EYSENCK, H. J. The dynamic theory of anxiety and hysteria. *J. ment. Sci.*, 1955, **101**, 28–51.

EYSENCK, H. J. The questionnaire measurement of neuroticism and extraversion. *Riv. Psicol.*, 1956, **50**, 113–140.

EYSENCK, H. J. *The dynamics of anxiety and hysteria*. London: Routledge & Kegan Paul, 1957.

FRANKS, C. M. Conditioning and personality: A study of normal and neurotic subjects. *J. abnorm. soc. Psychol.*, 1956, **52**, 143–150. (a)

FRANKS, C. M. The Taylor scale and the dimensional analysis of anxiety. *Rev. Psychol. appl.*, 1956, **6**, 35–44. (b)

FRANKS, C. M. The effect of food, drink and tobacco deprivation on the conditioning of the eyeblink response. *J. exp. Psychol.*, 1957, **53**, 117–120. (a)

FRANKS, C. M. Personality factors and the rate of conditioning. *Brit. J. Psychol.*, 1957, **48**, 119–126. (b)

GUILFORD, J. P. *Inventory of factors STDCR*. Beverley Hills, Calif.: Sheridan Supply, 1940.

HATHAWAY, S. R. *Supplementary manual for the MMPI: The K scale and its use*. New York: Psychological Corporation, 1946.

HATHAWAY, S. R., & MCKINLEY, J. C. *The Minnesota Multiphasic Personality Inventory*. New York: Psychological Corporation, 1943.

HILL, W. F. Comments on Taylor's drive theory and manifest anxiety. *Psychol. Bull.*, 1957, **54**, 490–493.

JENKINS, J. J., & LYKKEN, D. T. Individual differences. *Annu. Rev. Psychol.*, 1957, **8**, 79–112.

KERRICK, J. S. Some correlates of the Taylor manifest anxiety scale. *J. abnorm. soc. Psychol.*, 1955, **50**, 75–77.

LYKKEN, D. T. A study of anxiety in the sociopathic personality. *J. abnorm. soc. Psychol.*, 1957, **55**, 6–10.

MCCLELLAND, D., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.

MALTZMAN, I., FOX, I., & MORRISETT, L., JR. Some effects of manifest anxiety on mental set. *J. exp. Psychol.*, 1953, **46**, 50–54.

RAVEN, J. C. *Guide to using Progressive Matrices 1938*. London: Lewis, 1954.

SPENCE, K. W., & FARBER, I. E. Conditioning and extinction as a function of intensity of anxiety. *J. exp. Psychol.*, 1953, **45**, 116–119.

SPENCE, K. W., FARBER, I. E., & TAYLOR, E. The relation of electric shock and anxiety to level of

performance in eyelid conditioning. *J. exp. Psychol.*, 1954, **48**, 404–408.

SPENCE, K. W., & TAYLOR, JANET A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.

SPENCE, K. W., & TAYLOR, JANET A. The relation of conditioned response strength to anxiety in normal, neurotic, and psychotic subjects. *J. exp. Psychol.*, 1953, **45**, 265–272.

TAYLOR, JANET A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.

TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

TAYLOR, JANET A. Drive theory and manifest anxiety. *Psychol. Bull.*, 1956, **53**, 303–320.

WESLEY, E. L. Perseverative behavior in a concept-formation task as a function of manifest anxiety and rigidity. *J. abnorm. soc. Psychol.*, 1953, **48**, 129–134.

WILLETT, R. A. Conditioning, associative measures, and personality. In H. J. Eysenck (Ed.), *Experiments in personality*. London: Routledge & Kegan Paul, 1960.

(Received October 27, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 63, No. 3, 524–529

SOCIAL AND PSYCHOLOGICAL VARIABLES IN LEARNING HEBREW[1]

MOSHE ANISFELD AND WALLACE E. LAMBERT
McGill University

THE role of social attitudes in the learning of foreign languages has caught the attention of social scientists as well as language teachers (Nida, 1957–58; Vachon-Spilka, 1959; Whyte & Holmberg, 1956). Gardner and Lambert (1959) and Gardner (1960) have carried out empirical studies of this aspect of second language learning. The available findings indicate that the attitude of a student of a second language towards the other language community plays an important role in his acquisition of the other language. Students who hold sympathetic attitudes toward the other language community are found to be more successful in acquiring a second language than those holding unfavorable attitudes.

Several theoretical positions have been presented to account for these findings (Anisfeld, 1960; Ervin, 1954; Gardner, 1960). Our own theoretical scheme, presented in detail elsewhere (Anisfeld, 1960), emphasizes the generalization of attitudes developed toward a linguistic group to the language they speak (and vice versa). From this perspective, a language has potential positive or negative reinforcing value.

Since Hebrew is associated exclusively with the Jewish people, unlike many other languages that are spoken by various cultural groups, the generalization hypothesis seemed particularly appropriate for this language. The specific hypothesis was posed that the relation between attitudes toward the other language group and language achievement, found to hold for other languages, should be particularly prominent in the case of Jewish children learning Hebrew, in which their own language is learned as a second language. A favorable attitude towards belonging to the Jewish community should facilitate the children's acquisition of Hebrew. On the other hand, an attitude of marginality vis-à-vis the Jewish group should interfere with achievement in Hebrew.

Another related prediction follows from Gardner and Lambert's (1959) distinction between "integrative" and "instrumental" orientations towards the study of a foreign language. In their study, students' purposes for studying a foreign language were classified as integrative when the aim was to "learn more about the language group, or to meet more and *different* people," and as instrumental when the reasons reflected the "more utilitarian value of linguistic achievement" such as getting a better position, becoming better educated. In this study, we attempt to test, in the case of Hebrew, the findings of Gardner and Lambert (1959) and Gardner (1960) that integratively oriented students are better learners of a second language than are instrumentally oriented students.

The predictive power of the Psi-Lambda Foreign Language Aptitude Battery is also tested for the case of high school students studying Hebrew. Carroll (1960) mentions the use of this test with adults for predicting success in intensive courses of Hebrew with results similar to those found for other languages in similar training situations.

METHOD

Achievement Measures

Scores on the following tests were obtained:

1. *Teachers' ratings.* The Hebrew instructor of each class was asked to assign to his students the usual school marks for achievement in Hebrew with as much distribution as possible. In some cases, teachers were approached a second time and asked to rank those students who were given the same mark. Since most teachers tended to stress reading ability and understanding of written language, we asked them to pay special attention when assigning ratings to their students' ability to speak and comprehend Hebrew.

[1] This research was supported by the Canadian Defence Research Board, Grant Number D77-94-01-10.

We are grateful to S. Cytrynbaum for his assistance in the administration and scoring of the tests, to D. Solomon (McGill) and L. Rosenberg, research director of the Canadian Jewish Congress, for their critical readings of a preliminary draft relating to the characterization of the various Jewish congregations in Montreal.

In one class (Class A) which was at a sufficiently advanced level to enable the teacher to make a distinction between her students' comprehension of Hebrew and their oral skills in the language, she was asked to rate them separately on these two aspects of performance.

2–4. *Voice recordings.* For one school 2-minute tape recordings of the subjects' readings of a difficult Hebrew passage were obtained. These recordings were evaluated by two Hebrew experts on each of the following three aspects: Reading Fluency (2), Pronunciation Accuracy (3), Purity of Accent (4).

The score on each skill was taken as the combined ratings of both judges. The interjudge agreement was 88%.

Ability Measures

5. *Verbal Reasoning test.* This test, taken from the Differential Aptitude Tests (Bennett, Seashore, & Wesmen, 1959), was used as a measure of intelligence. Because of time limitations the time allowance for this test was shortened from 30 to 15 minutes.

6. *IQ.* Estimates of IQ were available for one class (Class A) only. These estimates were based on the Otis Self-Administering Tests of Mental Ability which had been administered by the Jewish Vocational Guidance Centre in 1958.

7. *Psi-Lambda Foreign Language Aptitude Battery, Short Form* (Carroll & Sapon, 1956). The tests used were: Spelling Clues, Words in Sentences (shortened from 15 to 10 minutes), and Paired Associates.

Attitude Measures

8. *Anti-Semitism (A-S) scale.* This scale is composed of seven items, of which six were selected from the California A-S scale (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950) on the basis of their fitness for Jewish subjects. Minor changes in the wording of some items were introduced to make them appropriate for Canadian Jewish students. We employed this scale as a measure of Jewish "self-hatred."

9. *Orientation Index.* On the basis of preliminary discussions with Jewish educators and analysis of compositions written by Jewish pupils in Montreal on topics related to study of Hebrew and other languages, 14 possible reasons for learning Hebrew were composed. Nine judges were asked to classify these statements as either integrative or instrumental according to the definitions of Gardner and Lambert (1959). The criterion of inclusion in the scale was agreement of at least seven judges on the categorization of an item. Six items, three integrative and three instrumental, met this criterion and were finally used to make up the Orientation Index.[2] For each of these statements the subject was requested to state, along a three-point scale, the degree of importance the particular reason had for him.

The integrative and instrumental items were scored separately, and then the instrumental score was subtracted from the integrative score to obtain an index of the balance between the two orientations for each individual. Thus the Orientation Index measures how much more (or less) intense the student's integrative motive is than his instrumental motive for studying Hebrew.

Subjects

Although 11 classes in seven Jewish parochial schools in Montreal were contacted, for reasons described elsewhere (Anisfeld, 1960), only 7 classes were used for the final analysis. The main reason for not analyzing all classes was that the samples were so small. In this paper we report the results of 6 classes from three schools in different settings in Montreal. (Because of doubts regarding the validity of the achievement measures of the seventh class, it is not included in the present discussion.) The schools as well as the classes so obviously differed in level of training, experience required, and regularity of attendance, in addition to having different instructors, that it was appropriate to treat each class as a separate sample (Guilford, 1950, pp. 352–354). A description of each class is therefore in order.

Class A ($N = 16$) is a Grade 9 class in an all-day Hebrew school. All children have attended the elementary branch of this school, so that when tested they were in their ninth year of experience in Hebrew studies. The children study the Hebrew language for 5 hours weekly in addition to their contact with the language in other Jewish subjects, such as Bible and Talmud study, and reading of prayers.

Class B ($N = 11$) is a Grade 8 class in another school. All students except for one have already gone through the all-day elementary branch of the school, and they now attend only Sunday and afternoon classes. They receive 3.5 hours of instruction in Hebrew subjects and 2.5 in Yiddish subjects during the week.

We also tested four classes in School C ($N = 54$) which is the largest and wealthiest Jewish afternoon school in Montreal. We will refer to these classes as *Class C (8a)*, *Class C (8b)*, *Class C (9a)*, and *Class C (9b)*, with the number denoting grade, *a* and *b* standing for the "better" and the "poorer" classes of students, respectively. In this school the children receive 4 hours of instruction in Jewish subjects during the week. Most pupils have been attending this afternoon school since Grade 1.

The students were divided into better and poorer classes when they completed Grade 2. We wanted to understand on what basis this classification was made. The classes were therefore compared on the following variables: Verbal Reasoning (Variable 5), and Anti-Semitism (Variable 8). The *t* tests indicate that only

[2] The integrative reasons were:
2. Hebrew is necessary to keep the Jewish people together.
4. I want to become more a part of the Jewish culture.
5. In order to be a good Jew I have to know Hebrew.

The instrumental reasons were:
1. One can enjoy a more successful life in the community with a knowledge of Hebrew.
3. In the future it may prove useful in getting a job.
4. I wish to enter a profession in which it is necessary to be familiar with Hebrew.

the differences in language aptitude are significant. The t values for the differences between the a and b classes on each grade level are significant beyond the .02 level (on two-tailed tests). It appears reasonable to conclude that the factor underlying the differences in achievement between the a and b classes is language aptitude.

RESULTS AND DISCUSSION

In order to test our hypotheses relating both ability and attitude to achievement in Hebrew, the correlations of these two classes of prediction variables with achievement measures were computed for each of our six subsamples.

The heterogeneity of the classes in our study not only called for statistical treatment of each group separately but also prevented us from averaging the coefficients to obtain a single estimate of the population correlation. Since some of the correlations of the prediction variables with achievement measures differed significantly from class to class, we found it unjustified to assume that "the several r's did arise by random sampling from the same population," an assumption necessary to allow the averaging of coefficients (Guilford, 1950, p. 355). Furthermore, there is also other evidence, to be presented later, that suggests that the schools investigated represent samples from different populations.

Ability and Achievement

Table 1 presents the correlations of our two indices of ability (intelligence and linguistic aptitude) with achievement measures. Linguistic aptitude as measured by the Psi-Lambda battery correlates quite highly with achievement measures for all classes. We notice in the Class A rubrics in Table 1 that the correlations between Foreign Language Aptitude and Purity of Accent and Oral Skill do not reach significance. This appears to be in line with the findings of Gardner (1960) who reports insignificant correlations between each of the subtests of the Short Form Psi-Lambda battery (the same that was used in the present study) and accent characteristics. However, the Short Form does not include the Phonetic Script test which measures the "ability to 'code' auditory phonetic material in such a way that this material can be recognized, identified, and remembered over something longer than a few seconds. . . . [an ability which is important not only in] remembering phonetic material (words, forms, etc.) but also in mimicking speech sounds" (Carroll, 1960, pp. 46–47). The auditory coding ability and the ability to mimic are obviously necessary for speech production; it therefore seems that Gardner's conclusion is not justified. As a matter of fact, in Gardner's (1960) study the correlation between the Spelling Clues test, which gauges the ability to produce sounds, and accent comes very close to significance, as do also the correlations of the total Short Form with Accent and with Oral Skill in the present study (Table 1).

Turning to intelligence, we observe in Table 1 that for Class A (which is at a sufficiently advanced level of proficiency in Hebrew to make possible a distinction among several aspects of linguistic skill) both Verbal Reasoning and IQ show differences in their correlations among various aspects of performance in Hebrew. Both indices of intelligence correlate highly with Comprehension of Hebrew and with Oral Skill but not with Fluency, Pronunciation Accuracy, and Purity of Accent. This finding appears to support the view (Dunkel, 1948) that intelligence plays a differential role in various language domains, with success in the more

TABLE 1
CORRELATIONS OF ABILITY VARIABLES WITH ACHIEVEMENT MEASURES

Achievement measures	Ability measures		
	Psi-Lambda	Verbal Reasoning	IQ
Class A (N = 16)			
Fluency	.599**	.079	.244
Pronunciation Accuracy	.613**	.098	.164
Purity of Accent	.420	.093	.068
Comprehension (Teacher's ratings)	.667***	.492	.546*
Oral Skill (Teacher's ratings)	.432	.377	.321
Class B (N = 11)			
Teacher's ratings	.817***	.663*	
Class C (8a) (N = 13)			
Teacher's ratings	.532	.409	
Class C (8b) (N = 11)			
Teacher's ratings	.621*	.700***	
Class C (9a) (N = 14)			
Teacher's ratings	.417	−.245	
Class C (9b) (N = 11)			
Teacher's ratings	.647*	.774***	

* Significant at .05 level; two-tailed test.
** Significant at .02 level; two-tailed test.
*** Significant at .01 level; two-tailed test.

cognitive tasks, represented by Comprehension, being more dependent on intelligence than success in tasks involving mainly oral production. The considerable magnitude of the correlations of Verbal Reasoning and of IQ with the Oral Skill ratings is probably due to the fact that these ratings were assigned by the teacher who, very likely, took into account to some extent the cognitive aspects of oral proficiency, such as the ideas a student expresses and the way he presents them.

Attitudes and Achievement

Table 2 summarizes the relations between each of the attitude variables and the measures of achievement. The correlations of the Anti-Semitism (A-S) scale and of the Orientation Index with achievement measures seem to be dependent upon the particular classes examined. For Classes A and B, the A-S scale is a reliable predictor of achievement in Hebrew whereas the Orientation Index is not; the reverse is generally true for the C classes. A possible explanation for this patterning is suggested by a consideration of the differential social characteristics of our samples. Classes A and B are located in and draw their students from a dwelling district in Montreal (Outremont and vicinity) which is lower in the socio-economic class hierarchy than School C (located in the Westmount suburb).[3]

Briefly, the parents of our students in the Outremont district are first generation North Americans; they still bear revealing traces of their old-fashioned Jewish background which they apparently dislike and from which they strive to free themselves. The Outremont students are or have been day students for 7 or 8 years, and they are advanced enough in Hebrew that both the teachers' ratings and the objective tests of achievement can be regarded as reliable discriminators of their ability in the language. Consequently, our finding of a correlation between attitudes toward the Jewish group (A-S) and proficiency in Hebrew should be taken as encouraging support of our hypothesis. On the other hand, the failure to find the same relation for three of the Westmount classes is not necessarily conflicting evidence. Because the Westmount students are at a very elementary level of Hebrew and the individual differences between the students in achievement are slight, it seems likely that the teachers, when asked to rank their students in Hebrew skill, resort to considerations of other classroom behaviors which actually differentiate these students. The Anti-Semitism variable does not seem to be one of these behaviors. It is difficult to imagine why Westmount Jews should be anti-Semitic. They are already second-generation North Americans, have adapted to the American culture to a considerable extent, and are economically and socially secure. The dimension of own-group devaluation is comparatively meaningless for Westmount Jews. It is our belief that they are relatively unconcerned with feelings of self-devaluation. It follows that the variation obtained on the A-S scale for the C classes may well be chance variance,[4] and the scale should not be regarded as measuring for the Westmount students a latent variable around which psychological processes are organized—a conception of attitudes presented by Green (1954).

The Orientation Index, measuring a desire to integrate in the Jewish culture, also seems to manifest differential sensitivity from class to class. The Outremont children apparently take their Jewishness as a matter of obvious fact and are not concerned with problems of integration in the Jewish culture. For them the variance on the Orientation Index is presumably chance variance and consequently this index does not predict achievement in Hebrew. In contrast, the problem of understanding what being Jewish means, of establishing a cultural identity, is of major concern to our Westmount students and their parents. It would not be surprising, therefore, if the Westmount teachers in evaluating a student's standing in the Hebrew language, a major subject in the Hebrew school, took into account the degree a student exhibits a willingness to integrate in the Jewish culture through the Hebrew school. Furthermore, to the extent that the teachers' ratings do reflect true individual differences in Hebrew *achievement*, the integration dimension would be a salient one

[3] The following characterization of the Jewish residents of these districts is based to a great extent on a sociological study of Montreal Jewish areas by Seidel (1939) and draws on the theory of Tryon (1955).

[4] The two districts do not differ in mean scores and variances on either the A-S scale or the Orientation Index.

TABLE 2
CORRELATION OF ATTITUDE VARIABLES WITH ACHIEVEMENT MEASURES

Achievement measures	Attitude measures	
	Anti-Semitism	Orientation[a]
Class A (N = 16)		
Fluency	−.504*	−.329
Pronunciation Accuracy	−.488	−.143
Purity of Accent	−.438	−.078
Comprehension (Teacher's ratings)	−.589**	.203
Oral Skill (Teacher's ratings)	−.495*	−.054
Class B (N = 9)		
Teacher's ratings	−.682*	−.405
Class C (8*a*) (N = 15)		
Teacher's ratings	−.576*	−.612**
Class C (8*b*) (N = 11)		
Teacher's ratings	−.084	.632*
Class C (9*a*) (N = 15)		
Teacher's ratings	−.160	−.514*
Class C (9*b*) (N = 11)		
Teacher's ratings	−.075	−.371

[a] Correlation is positive if "integrative" orientation is related to achievement; correlation is negative if "instrumental" orientation is related to achievement.

* Significant at .05 level; two-tailed test.

** Significant at .02 level; two-tailed test

for the Westmount students and therefore contribute to differences in success in Hebrew. These two explanations may account for the correlations of the Orientation Index found in the Westmount school.

Two of the three significant correlations of the Orientation Index with teachers' ratings are negative. We predicted a positive correlation between the Orientation Index and achievement, indicating that the integrative motive—which expresses a desire to become better integrated in the Jewish culture—is more conducive to success in Hebrew than the instrumental motive. This discrepancy suggests a reconsideration of the integrative-instrumental distinction as defined by our index. It seems that in the case of Hebrew the instrumental items really indicate a greater desire to integrate in the Jewish culture than the integrative items. Wanting to get a job requiring knowledge of Hebrew (an instrumental reason for studying Hebrew) actually means becoming a rabbi, a Hebrew teacher, or engaging in other Jewish professions which involve more participation in the Jewish culture and community than the more idealistic and vague purposes of becoming "more a part of the Jewish culture" (an integrative reason). The situation is different for English-speaking children in Catholic schools in bilingual Montreal where Gardner and Lambert (1959) and Gardner (1960) found that the integrative orientation led to superior achievement than did the instrumental orientation. In Montreal there are positions requiring knowledge of French which are clearly instrumental and do not imply any integrative feelings towards the French-Canadians. Our results may suggest that the reasons for studying a language, whether instrumental or integrative are not important in themselves: a particular orientation influences language acquisition only to the extent that it is indicative of attitudes towards the language community.

Interrelations among Prediction Variables

This discussion suggests that for a student to do well in Hebrew, and probably in other languages as well, he needs to possess intelligence, language aptitude, and a proper attitude toward the language community. Inspection of the intercorrelations among these variables indicates that some of the correlations between the prediction variables and the achievement measures may be artificially high. The correlations of Psi-Lambda with Verbal Reasoning for the six classes under our consideration range from .47 to .83 with only one exception of a correlation of .19. (Similar results with a different intelligence test are documented by Gardner & Lambert, 1959.) Partialing out the intercorrelation from either of these variables would certainly reduce their correlations with achievement measures. Though much lower in magnitude, there is also some relation between A-S and Psi-Lambda. The correlations for Classes A, B, C (8*a*), C (8*b*), C (9*a*), and C (9*b*) are −.27, −.48, .09, −.84, .01, and .01, respectively. A partial correlation between Psi-Lambda and achievement measures and between A-S and the same measures would lower some of the correlations reported in this paper. However, the present study has demonstrated the relevance of attitudes to the acquisition of a new language, but because of the smallness of samples and inequality of Ns for different variables it is impossible to determine more precisely the role of each variable through factor analysis or partial correlations.

The relation between Psi-Lambda and A-S

warrants interest apart from its possible influence on the correlations of these variables with achievement. There appear to be at least two plausible cause-effect interpretations of this finding. The causal relation may rest between language aptitude and anti-Semitism. The ability of a pupil to master the Hebrew language is a prerequisite to whatever success he may have in Hebrew school. The Hebrew school plays a central role in the Jewish community, and whatever attitudes a student develops toward the Hebrew school hold for the Jewish community at large. The student with language aptitude would find it easier to learn Hebrew, and consequently do well in Hebrew school, be rewarded for his achievement and come to attach positive value to the Hebrew school and the Jewish community—an indirect source of reinforcement. His favorable attitude to the Jewish cultural group could express itself as a low score on the A-S scale. In comparison, students who lack language aptitude would find it more difficult to learn Hebrew; possibly they associate the Hebrew language and school with frustrating experiences which, when generalized as an attitude to the Jewish community, would yield high scores on the A-S scale. Such an explanation is consonant with our generalization hypothesis, suggesting that the generalization process works not only from a community to its language but also from a language to the language community. An alternative interpretation is that the negative correlations between anti-Semitism and Psi-Lambda are due to the influence of some common factor. Since the F Scale (which is highly related to A-S) does not correlate with the Psi-Lambda tests (Gardner, 1960; Gardner & Lambert, 1959), it is difficult to speculate about the nature of such a common factor.

SUMMARY

A study was conducted with Jewish eighth and ninth grade students learning Hebrew. Measures of achievement in the Hebrew language were correlated with scores on two groups of tests: attitude tests which measure their reasons for learning Hebrew and their attitudes toward the Jewish culture and community, and ability measures including intelligence and language aptitude tests. The results support our predictions about the relation of both ability and attitude to success in learning Hebrew. However, whereas intelligence and linguistic aptitude are relatively stable predictors of success, the attitude measures are less stable and vary in their relationship to achievement from school to school and from class to class. We attempted to explain these differences through considerations of the sociopsychological characteristics of the Jews in the various districts of Montreal where the schools are located.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* New York: Harper, 1950.

ANISFELD, M. Social and psychological variables in learning Hebrew. Unpublished master's thesis, McGill University, 1960.

BENNETT, G. K., SEASHORE, H. G., & WESMEN, A. G. *Manual for the differential aptitude tests.* New York: Psychological Corporation, 1959.

CARROLL, J. B. The prediction of success in intensive foreign language training. Prepared for a symposium on Training Research, University of Pittsburgh, 1960. (Final rev. Harvard University, 1960 [Mimeo]).

CARROLL, J. B., & SAPON, S. M. Supplement 1 to examiner's manual—Psi-Lambda foreign language aptitude battery. Harvard University, 1956. (Mimeo)

DUNKEL, H. B. *Second-language learning.* Boston: Ginn, 1948.

ERVIN, SUSAN. Identification and bilingualism. Harvard University, 1954. (Mimeo)

GARDNER, R. C. Motivational variables in second-language acquisition. Unpublished PhD thesis, McGill University, 1960.

GARDNER, R. C., & LAMBERT, W. E. Motivational variables in second-language acquisition. *Canad. J. Psychol.*, 1959, **13**, 266–272.

GREEN, B. F. Attitude measurement. In G. Lindzey (Ed.), *Handbook of social psychology.* Vol. I. Cambridge, Mass.: Addison-Wesley, 1954. Pp. 335–369.

GUILFORD, J. P. *Fundamental statistics in psychology and education.* (2nd ed.) New York: McGraw-Hill, 1950.

NIDA, E. A. Some psychological problems in language learning. *Lang. Learning*, 1957–58, **8**, 7–16.

SEIDEL, JUDITH. The development and social adjustment of the Jewish community in Montreal. Unpublished master's thesis, McGill University, 1939.

TRYON, R. C. Identification of social areas by cluster analysis. A general method with an application to the San Fransisco Bay area. *U. Calif. Publ. Psychol.*, 1955, **8**(1), 1–100.

VACHON-SPILKA, IRENE. Communication aux membres de l'Association Canadienne de Linguistique. Saskatoon, 1959. (Mimeo)

WHYTE, W. F., & HOLMBERG, A. R. Human problems of U. S. enterprise in Latin America. *Hum. Organization*, 1956, **15**, 1–40.

(Received October 31, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 63, No. 3, 530-533

THE EXTINCTION OF FEAR AS A FUNCTION OF DISTANCE VERSUS DISSIMILARITY FROM THE ORIGINAL CONFLICT SITUATION[1]

THOMAS ELDER, CHARLES D. NOBLIN
Louisiana State University
AND BRENDAN A. MAHER[2]
Harvard University

MURRAY and Berkun (1955) have combined Miller's (1944, 1948) conflict and displacement models to form a three-dimensional scheme that generated predictions concerning the simultaneous operations of conflict and displacement. Conflict is derived from a set of assumptions relating response strength to the distance of the subject from the goal (Miller, 1944); displacement is based upon a similar set of assumptions concerning the difference in similarity between an original and other stimulus objects (Miller, 1944). Murray and Berkun diagrammatically represent this three-dimensional system by two horizontal axes at right angles to each other, representing distance and similarity scales, respectively. A third axis extends vertically from the point of intersection of the two horizontal axes and represents a scale of response strength. By placing two points on each of the three axes, two planes are obtained: one representing approach and one representing avoidance tendency, from which are derived specific predictions of behavior.

Applying the model to the behavior of rats, they deduced that subjects trained to approach-avoidance conflict in an original alley would, whenever possible, leave that alley to enter a different one and that the more dissimilar the new alley the nearer subjects would approach the goal. As a test of these deductions, they first trained rats to run an alley for food and on later trials established conflict by administering shock while the subjects were eating. These subjects were then placed at the starting end of the original alley and could either remain in this alley or could leave and enter new and different alleys adjacent and parallel to the original. Windows were spaced at equal intervals along the common walls of adjacent alleys to permit crossing from one to another. Tracings of the movements of the rats through the maze followed the pattern predicted; i.e., the subjects ran farther down the most dissimilar alley than down the moderately dissimilar alley, and farther down the moderate alley than down the original. They also reported that goal responses occurred first in the most dissimilar alley, second in the intermediate, and finally in the original alley. Third, they observed that subjects entered the most dissimilar alley through windows farther down the intermediate alley than the windows used when crossing from the original to the intermediate alley. These results were interpreted as confirmation of deductions from, and support for, the three-dimensional model insofar as it predicts the locus of "displacement" to be a function of the similarity between the original and the test situations.

A recent study (Maher, Elder, & Noblin, 1962) in which it was necessary to replicate the Murray and Berkun (1955) experiment has confirmed all but one of these findings. In this it was found that rats tend to enter more dissimilar alleys through the windows nearest the start end of the runways rather than windows spaced farther down the alleys. This finding was combined with the observation that Murray and Berkun always located the most dissimilar alley farthest away from the original, thus confounding distance and dissimilarity and making it impossible to decide if the subjects are entering new and different alleys as a response to distance cues or as a response to similarity cues.

If distance is the critical factor, then by deleting the principle of generalization from the

[1] This study was supported by funds from Grants G 3852/6433 from the National Science Foundation to Brendan A. Maher.

It was read in abbreviated form at the annual meeting of the Southeastern Psychological Association, Atlanta, Georgia, 1960.

[2] In the Department of Social Relations at Harvard University.

three-dimensional formulation, and extending the assumptions about distance of the subject from the goal to include the distance of other alleys from the original, a simplified model evolves from which the same prediction (i.e., the "therapeutic effect") can be generated.

The purpose of the present study was to devise a situation in which the spatial distance of dissimilar alleys and the dissimilarity of distant alleys from the original could be systematically altered, thereby subjecting to empirical tests two alternate deductions: Subjects trained to approach-avoidance conflict should either enter and run farther down the most dissimilar alley as predicted by the three-dimensional model; or as the simpler formulation predicts, they should enter and run down the alley most distant from the original.

METHOD

Apparatus

The apparatus (see Figure 1) consisted of four alleys each of which was 48 inches long, 5 inches wide, 6 inches deep, white, grey black, and white, respectively. The alleys and walls of the alleys contained four 2.5 inch square windows equally spaced along the walls of the alleys, each of separate construction so the parallel arrangement could be varied. A food cup 1.5 inches in diameter was attached to an end wall of each alley with the rim of the cup 3 inches from the floor. The floors of the alleys were covered with .25 inch hardware cloth. Hooks were placed on the outside of the end walls and a spring extended from the hooks on each end of the first alley to the hooks on the ends of the fourth alley to hold the four runways securely in parallel. A removable cover was constructed of .5 inch hardware cloth and placed on top of the four-alley combination.

Subjects

The subjects were 24 male, albino rats of the Sprague-Dawley strain. These were randomly assigned to six groups of four and placed on a 24 hour food deprivation schedule 7 days prior to runway training. All subjects were fed 8 grams of Purina dog chow cubes daily, and a plentiful supply of water was available at all times.

Procedure

On the first day of training each of the 24 subjects received 25 successive reinforced approach trials in a white alley, hereafter designated as the original alley. A windowless wall was inserted between the original and adjacent alley to prevent the subject from entering other alleys during training trials. This procedure was repeated on the second day. On the third day subjects each received 10 warm-up trials, after which they received shock at the goal end of the original alley. The .75 milliampere shock was administered when the subject made a circuit between the metal food cup and the hardware-cloth floor. Shock trials were continued until the subject no longer ran beyond a point 18 inches from the start end of the runway.

Testing trials commenced immediately following the shock trials. The windowless wall between the original and the adjacent alley was removed and a wall with windows inserted in the maze.

Since the original alley was the same for all subjects, only the parallel position of the others was altered. This permitted six possible parallel arrangements of the four alleys, with one group of subjects assigned to each arrangement. Thus each group was tested under a different color/position combination.

A test trial consisted of placing the subject at the start of the original alley, the experimenter tracing its path through the maze. A trial terminated with the

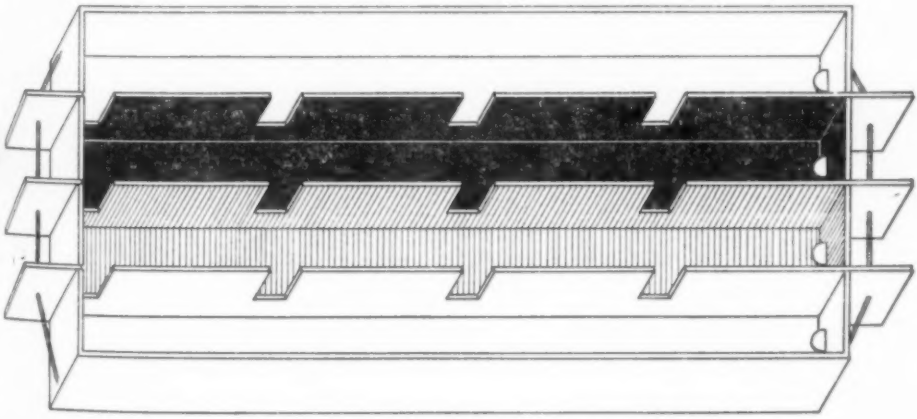

FIG. 1. Parallel alleys for testing displacement.

subject's making the goal response in any one of the runways or until a period of 2 minutes had elapsed. A goal response was defined by the subject's reaching up and placing its nose down into the food cup. Each subject received a maximum of 4 test trials on the first day and 10 test trials on each day thereafter until a criterion of two successive goal responses in the original alley had been reached.

RESULTS

A mean distance score was computed for each subject by summing the number of eighth's of the distance traveled down the alley on each trial, and dividing by the total number of trials in which the subject entered that particular alley. Such scores were computed for each of the 24 subjects, the mean result for the entire sample being given in Figure 2. The broken vertical line in the figure identifies scores obtained in the original from scores in the interchangeable alleys.

In order to test the effects of distance and/or color, a mixed analysis of variance on distance scores in the interchangeable alleys was used. As the summary in Table 1 shows, the distance variable produced the only significant effect. Variance due to groups (i.e., between-subjects, Distance × Color interaction) approached but did not reach the minimum significance level, and color appears to have had little or no effect. An unexpected result was that trials in the first half did not differ significantly from trials in the second half. Figure 3 shows that the only difference observed between early and later trials was in the original alley ($p < .001$ by a one-tailed sign test). The range of trials required to reach the criterion of two successive goal responses in the original alley was from 6 to 21 with a mean of 10.2

TABLE 1

SUMMARY OF ANALYSIS OF VARIANCE ON SUBJECTS' MEAN DISTANCE TRAVELED DOWN ALLEYS 2, 3, AND 4

Source	*df*	*SS*	*MS*	*F*
Between-subjects	11	145.04		
Groups (G)	2	47.61	23.80	2.20
Error (b)	9	97.43	10.82	
Within-subjects	60	419.23		
Distance (D)	2	96.23	48.11	9.56*
Color (C)	2	6.31	3.16	.63
Trial blocks (T)	1	2.51	2.51	.50
D × T	2	5.90	2.95	.59
C × T	2	11.74	5.87	1.17
G × T	2	9.78	4.89	.97
D × C (w) × T	2	2.84	1.42	.28
Error (w)	47	236.31	5.03	
Total	71	564.27		

* $p < .01$.

To test Murray and Berkun's (1955) prediction that subjects should tend to use windows increasingly farther down toward the goal of increasingly dissimilar alleys, exit scores for the alleys along a distance as well as a similarity dimension were derived in the manner described by Murray and Berkun. The means are given in Figure 4. A one-tailed sign test between scores of Alleys 1 and 2, 2 and 3, and 3 and 4 for both distance and color failed to yield a significant difference.

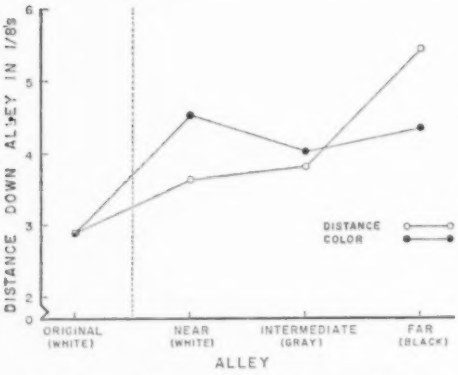

FIG. 2. Mean distances traveled down each alley.

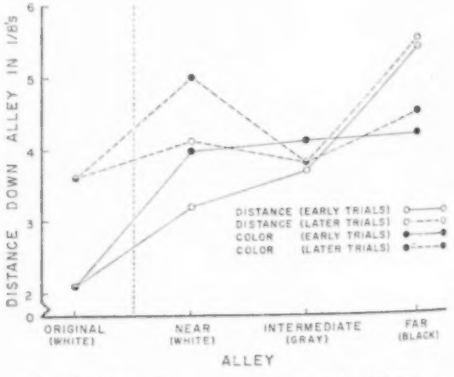

FIG. 3. Mean distance scores on early trials compared with later trials.

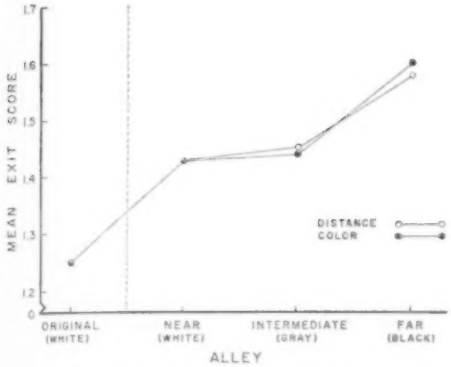

FIG. 4. Mean exit scores for each alley.

DISCUSSION

The hypothesis derived from the approach-avoidance-displacement model that subjects should run farther down more dissimilar alleys is rejected. The alternative hypothesis that rats trained to approach-avoidance conflict run farther down adjacent alleys as function of the spatial distance of the other alleys from the original is supported. However, the failure of the gradient to shift toward the goal ends of the neighboring alleys, and the dramatic forward shift of the point of intersecting approach-avoidance gradients in the original alley raises some doubt as to the nature of the therapeutic effect produced by experience in the other alleys. The tendency for subjects to leave the original alley and enter some alleys may, perhaps, be most parsimoniously viewed as *escape* behavior, rather than as a response to conflict generalized to other situations.

A possibility exists, however, that the dissimilar alleys did have a therapeutic effect, as a function not of *degree* of dissimilarity but of dissimilarity per se. While the behavior of subjects in the dissimilar-white alley differed from that observed in the original-white alley, thereby seeming to cast doubt on this hypothesis, these two alleys, similar in color, may have been dissimilar in some uncontrolled respect such as odor.

At this point, the conclusion drawn from this study is restricted to the inference that the three-dimensional model of Miller and others is open to question insofar as it utilizes a dissimilarity dimension to account for displacement.

SUMMARY

The influence of distance was compared with that of dissimilarity in producing displacement behavior in an approach-avoidance conflict. Twenty-four white rats were used in a design that permitted separation of the effect of these two variables. Increasing reappearance of the approach response was significantly related to distance but not to dissimilarity. Tentative explanations are offered in terms of escape behavior, or of uncontrolled similarity cues such as odor. The Miller model of approach-avoidance-displacement is not supported by these data.

REFERENCES

MAHER, B. A., ELDER, S. T., & NOBLIN, C. D. A differential investigation of avoidance reduction versus hypermotility following frontal ablation. *J. comp. physiol. Psychol.*, 1962, in press.

MILLER, N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* Vol. 1. New York: Ronald, 1944. Pp. 431–465.

MILLER, N. E. Theory and experiment relating psychoanalytic displacement to stimulus response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155–178.

MURRAY, E. J., & BERKUN, M. M. Displacement as a function of conflict. *J. abnorm. soc. Psychol.*, 1955, **51**, 47–56.

(Received November 7, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 63, No. 3, 534-539

INCUBATION EFFECT:

ROLE OF AWARENESS IN AN IMMEDIATE VERSUS DELAYED TEST OF CONDITIONED EMOTIONALITY[1]

SANFORD GOLIN[2]
State University of Iowa

VARIOUS investigators (Bindra & Cameron, 1953; Diven, 1937; Haggard, 1943; Mednick, 1957) have reported that a delay between the acquisition of a conditioned emotional response and the subsequent presentation of the CS results in an increase in the strength of the conditioned response. It has been suggested that this increase in emotionality after a delay is related to the clinical phenomenon commonly called the "incubation" of anxiety (e.g., White, 1948).

A study by Diven (1937) has frequently been cited as providing experimental evidence for the incubation effect. Diven tested the effects of delay on emotionality by comparing a conditioned GSR immediately after acquisition with the GSR after periods of delay varying from 30 minutes to 48 hours. This experiment consisted of 10 subexperiments each differing with respect to some experimental variable. In making overall comparisons the effects of various variables were inextricably confounded and, thus, the evidence for concluding that emotionality increased after a delay was not clear. Subsequent studies have reported that delay resulted in an increase in emotionality (Bindra & Cameron, 1953), in a decrease (Haggard, 1943), and in an increase for responses to generalized stimuli but not for responses to the CS (Mednick, 1957). But these studies are not strictly comparable since they differed with respect to one or more variables such as the length of the CS-UCS interval and the point in this interval at which the GSR was measured.

Diven (1937) further reported that the increase in emotionality after a delay was relatively greater when subjects were unaware of the relation between the CS and UCS than when they were aware of this relation. This finding was considered consistent with the clinical hypothesis that repressed traumatic experiences are capable of generating more affect than trauma accessible to conscious experience. For methodological reasons Diven's evidence is dubious and, furthermore, the results of other studies (Haggard, 1943; Lacey, Smith, & Green, 1955; Mednick, 1957) imply that increases in emotionality after a delay are greater when the subject is aware of the CS-UCS relation. Mednick (1957), like others (e.g., Bitterman & Holtzman, 1952; Moeller, 1954) noted that GSR response strength declined after reaching a maximum early in acquisition—i.e., adaptation occurred—and suggested that the incubation effect may have been due to recovery from adaptation occurring during delay. Both Haggard (1943) and Lacey et al. (1955) found the adaptation of emotional responses to be greater for "aware" subjects than for "unaware" subjects. Hence, if incubation is a result of recovery from adaptation, aware subjects would be expected to show a greater increase in emotionality than unaware subjects as a function of delay.

This study was designed to test the view that emotionality, as measured by the GSR, increases after a delay and that this increase is greater when the subject is aware of the relation between the CS and UCS. Unlike previous studies in this area awareness was determined by allowing the subject to identify from a checklist of all stimuli those which were followed by the UCS. Hills (1959) found that written rather than oral report increases the proportion of subjects who are considered aware when awareness is defined in terms of ability to specify the CS-UCS relationship.

[1] Based on a dissertation submitted in partial fulfillment of the requirements for the PhD degree at the State University of Iowa. Acknowledgment is made to I. E. Farber, committee chairman, for his advice throughout this study.

[2] Now at the University of Wisconsin-Milwaukee.

METHOD

Subjects, Apparatus, Preliminary Procedure, and Materials

Subjects were 60 male undergraduate students from the State University of Iowa assigned in equal numbers to each of three List conditions described below.

After each subject was comfortably seated, GSR pick-up electrodes and shock electrodes were attached. After a "ready" signal, the subject was given a mild (15 volt) electric shock of 1 second duration, administered through copper electrodes attached to the palm and middle finger of the left hand, by means of a Variac connected to a 120 volt isolated alternating current. The GSR pick-up electrodes, attached to the middle and little fingers of the right hand, consisted of round lead plates $\frac{3}{8}$ inches in diameter and $\frac{1}{8}$ inch thick. A felt corn plaster filled with Sanborn electrode paste provided the actual contact with the subject's finger.[3] The GSR was recorded as a change in conductance by an Esterline-Angus recorder through a Hunter GSR amplifier circuit. Time intervals and durations of stimuli were controlled by decade-type interval timers (Hunter & Brown, 1949). The shock apparatus, recorder, amplifier, and timers were located behind a 5 feet × 3 feet black masonite screen. After the subject was given the 15-volt shock, he was informed that the shock in the experiment proper would be similar to the 15-volt shock, but much stronger.

The stimuli consisted of 15 words (Table 1) presented for one second each on a 3 inch × 5 inch translucent cardboard screen by means of a 35 millimeter projector placed behind the masonite partition. Three of these words served as experimental words. The remaining words served as distractor stimuli to make the task more difficult. For each subject one of the three experimental words was a shock-aware (SA) word, i.e., the subject was informed beforehand that this particular word would be followed by shock. For each subject another of the three experimental words was a shock-unaware (SU) word, i.e., always followed by shock but not mentioned in the instructions. The last of the three words was a nonshock (NS) word, i.e. neither followed by shock nor mentioned in the instructions. The instructions were a modification of those employed by Lacey et al. (1955) which were designed to obscure the relation between the SU word and the shock.

Acquisition and Test Series

Six trials, each trial containing the three experimental words and six of the nine filler words, were presented to each subject. Subjects were randomly assigned to three List groups of 20 subjects each and the possible differential effects of the experimental words themselves were controlled by counterbalancing them among the three List groups, as shown in Table 1. The words were presented to the three List groups in the same order and therefore the effect of order with respect to Awareness (SA, SU, and NS) was also counterbalanced among the List groups.

[3] Suggested in a personal communication by Guy Miles from a technique devised by David Lykken.

TABLE 1
STIMULUS WORDS

Experimental words

Shock condition	List I	List II	List III
Shock-aware (SA)	TIGER	ZINC	COAT
Shock-unaware (SU)	ZINC	COAT	TIGER
Nonshock (NS)	COAT	TIGER	ZINC

Filler words

LEOPARD, WILDCAT, COPPER, TIN, SHIRT, DRESS, TRIANGLE, CIRCLE, RECTANGLE, SQUARE, HEXAGON, CONE

When the subject's GSR reached a base level the first stimulus word was presented. The subject was instructed to chain-associate to each word and tap a telegraph key located in front of him with the left hand at a constant rate until either a 6-watt light in the upper right corner of the masonite partition went on or until he received an electric shock. The subject chain-associated and tapped to each word for 15 seconds. If the stimulus word was an NS word, the end of the interval was indicated by the light. If the stimulus was an SA or SU word, the end of the interval was signaled, not by the light, but by a 100-volt electric shock, a stimulus shown by preliminary study to be safe, though painful. In order to allow the GSR sufficient time to reach a base level there was an interval of 30 seconds between the end of one trial and the beginning of the next. If at the end of 30 seconds the GSR did not reach a base level, the experimenter waited until a base level was reached before beginning the next trial. It is estimated that the GSR reached a base level by 30 seconds on 95% of the trials and those subjects who tended to have longer intertrial intervals were quite evenly distributed among the experimental groups.

After completion of acquisition one half of the subjects in each of the three List groups were randomly assigned to the Delay condition. The remaining half were assigned to the Immediate condition. The electrodes were removed from the subjects in the Delay condition who were then taken to another part of the room where they spent 30 minutes in a task involving abstract thinking problems. At the end of this period, these subjects were seated in front of the masonite screen as before, the electrodes being attached, and the test series (extinction) was begun. Subjects in the immediate group proceeded directly from the acquisition series to the test series.

All subjects were given two test trials with the electrodes attached, each trial containing the three experimental words and six of the filler words. No shock was administered during the test series. The base level of each subject was recorded in raw ohms resistance and the GSR was recorded as a change in conductance for each presentation of an experimental word both during acquisition and the test series.

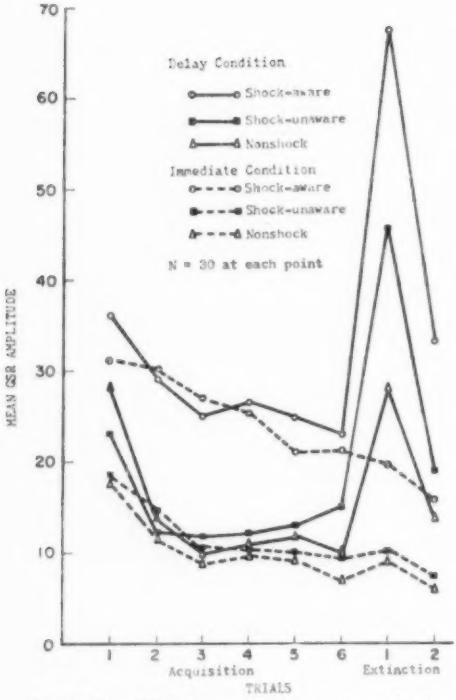

FIG. 1. Mean GSRs in conductance units under Immediate and Delay conditions.

Determination of Awareness

After the test series the electrodes were removed and each subject was presented with a written list of all the stimulus words and asked to check whether or not a word had been followed by shock during the acquisition series. The check list consisted of the following categories: (1) "am certain it was not," (2) "think it was not but not sure," (3) "don't remember," (4) "think it was but not sure," (5) "am certain it was." The data of those subjects who checked Category 5 for the SU word were not used in the analysis of results and additional subjects were randomly assigned to the List condition in which those subjects fell. There were 11 such subjects.

RESULTS

The GSR was measured in terms of the amplitude of change in conductance occurring between the onset of the CS and the onset of the signal, shock or light, terminating the end of the chain-association interval. Since the correlation between base level and GSR changes was found not to be significantly different from zero, the results were computed on the basis of the absolute magnitude of the GSR.

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE OF TRANSFORMED GSR DATA OF TRIAL SIX OF ACQUISITION

Source	*df*	*MS*	*F*	*p*
Between subjects	59			
Delay (D)	1	12.59	2.18	
Lists (L)	2	11.98	2.08	
D × L	2	3.52	<1	
Error (b)	54	5.77		
Within subjects	120			
Awareness (A)	2	54.16	33.13	<.001
A × D	2	.51	<1	
A × L	4	2.84	1.74	
A × D × L	4	4.01	2.45	
Error (w)	108	1.63		
Total	179			

Acquisition

Figure 1 shows the GSR acquisition curves for SA, SU, and NS words. The mean GSRs and *SD*s for Acquisition Trial 1 were, respectively, SA, 33.87, 26.04; SU, 20.68, 21.06; NS, 23.09, 23.03. The difference between the mean SA and mean NS GSRs (within-subjects comparison), which a t test showed to be significant ($p < .01$), indicated that the instructions were effective in producing an emotional response to the SA word before the shock was administered. There was no significant difference between the mean SU GSR and the mean NS GSR on Trial 1 ($p > .10$).

The mean GSRs and *SD*s at the end of acquisition (Trial 6) were, respectively, SA, 24.04, 23.64; SU, 12.25, 12.24; NS, 8.31, 10.02. In order to determine the effect of Awareness on the acquisition of the GSR an analysis of variance was carried out. In addition to Awareness, involving a within-subjects comparison, this analysis assessed the effects of two other factors, Lists and Delay (between-subjects comparisons). Since the assumption of homogeneity of interaction effects seemed untenable this analysis was carried out using square root transformations of the original data. The results, summarized in Table 2, showed that the GSR at the end of acquisition was significantly related only to awareness ($p < .001$). Individual t tests showed that the GSR to SA words was greater than that to SU words ($p < .001$) and NS words ($p < .001$) and the GSR to SU words was greater than that to NS words ($p < .02$). Since there

was no significant difference between the SU and NS words at the beginning of acquisition, the significant difference between them at the end of acquisition indicated that conditioning had occurred. The finding of no significant difference between Immediate and Delay subjects indicated there was no bias in the selection of subjects for these conditions since they received the same treatments during acquisition, while the finding of no significant effect of Lists showed that the GSR was neither a function of the words themselves nor of the interaction between words and the other factors.

As can be seen from Figure 1, the GSR to all three experimental words declined during acquisition, i.e., adaptation occurred. In order to compare the amount of adaptation between SA and SU words, the difference between the GSR on Acquisition Trial 2 and Acquisition Trial 6 (A-2 minus A-6) was computed for each subject. Trial 2 was taken as the starting point for assessing the relative effects of adaptation since it was the first trial following the first reinforced trial. The means and *SD*s of the difference scores were, respectively, SA, 8.84, 23.8; SU, 3.92, 12.5, the difference between the means (within-subjects comparison) being significant at between the .05 and .10 level.

Effects of Delay and Awareness

As can be seen from Figure 1, Delay subjects showed an increase from the last trial of acquisition (A-6) to the first test trial (T-1) whereas Immediate subjects did not. The difference between the GSR on T-1 and A-6 (T-1 minus A-6) was computed for each subject and an analysis of variance of the effects of Awareness (within-subjects) and of Delay and Lists (between-subjects) was carried out. The relevant means and *SD*s are presented in Table 3 and the analysis is summarized in Table 4. The analysis showed that the mean change in the GSR was significantly greater for Delay subjects ($p < .001$). Furthermore, the significant Delay × Awareness interaction ($p < .005$) showed that the increase in the GSR after a delay was a function of awareness; for Delay subjects, *t* tests showed that the increase was greater for the SA words than for either the SU words ($p < .01$) or NS words ($p < .001$), and the increase for SU words was greater than that for NS words ($p < .01$). There were no significant differences with respect to the increase from A-6 to T-1 for Immediate subjects.

TABLE 3

GSR DIFFERENCES BETWEEN TRIAL SIX OF ACQUISITION AND TEST TRIAL ONE

	Shock-aware	Shock-unaware	Nonshock
Immediate			
M	−1.68	.74	1.50
SD	14.57	8.26	7.15
Delay			
M	39.27	28.67	18.62
SD	46.41	42.30	28.58

TABLE 4

SUMMARY OF ANALYSIS OF VARIANCE OF GSR DIFFERENCES BETWEEN TRIAL SIX OF ACQUISITION AND TEST TRIAL ONE

Source	*df*	*MS*	*F*	*p*
Between subjects	59			
Delay (D)	1	36994.33	22.01	< .001
Lists (L)	2	889.86	<1	
D × L	2	595.68	<1	
Error (b)	54	1681.06		
Within subjects	120			
Awareness (A)	2	1147.78	3.64	< .05
A × D	2	2136.96	6.79	< .005
A × L	4	1599.01	5.08	< .005
A × D × L	4	4649.20	14.77	< .001
Error (w)	108			
Total	179			

The results just cited showed that both adaptation and incubation were greater for SA than SU words. To further determine if there was any relation between adaptation and incubation, Delay subjects were ranked with respect to the amount of adaptation and divided at the median into two groups, High and Low Adapters. A comparison of the mean incubation scores of these two groups revealed a significantly ($p < .05$) greater increase after a delay for High Adapters.

Further analysis of these significant Lists × Awareness ($p < .005$) and Lists × Awareness × Delay ($p < .001$) interactions showed that the results were related to word order on T-1, i.e., Delay subjects not only showed a differential increase with respect to awareness (SA, SU, and NS words), but also showed a significantly greater mean GSR to the first word within the trial presented after a delay, irrespective of the nature of the word. It will

be recalled that the order of presentation of the stimulus words was the same for all subjects but the order of the SA, SU, and NS "treatments" was counterbalanced among the List groups.

Since the SA, SU, and NS words differed not only with respect to the increase in the GSR from A-6 to T-1, but also with respect to their strength at the end of acquisition (A-6), a test was made of the hypothesis that the amount of increase was a function of the strength of the response at the end of acquisition. A comparison of high and low GSR responders on A-6 showed no significant difference with respect to the change from A-6 to T-1 ($p > .10$). Thus, it was shown that observed differences after a delay could not reasonably be attributed to the level of responding at the end of acquisition.

Figure 1 shows a large drop in the mean GSR for Delay subjects from the first (T-1) to the second (T-2) test trial. A comparison of A-6 with T-2 showed that Delay subjects, despite their rapid drop from T-1 to T-2, had not declined to their level of responding at the end of acquisition ($p < .025$).

Awareness Ratings

In the determination of awareness, 23 subjects checked Category 4 on the questionnaire, i.e., indicated that they thought the SU word was followed by shock but were not sure. The mean GSR to the SU word on the last trial of acquisition of these 23 subjects was compared with the comparable mean GSR of 23 subjects randomly selected from those subjects who checked Categories 1–3 for the SU word. A t test did not reveal any significant differences between those subjects who marked Category 4 and those who marked Categories 1–3.

DISCUSSION

The results of this study showed that an emotional response was learned without the subject being aware of the relation between the CS and UCS.[4] This result is in accord with the findings of Diven (1937) and Lacey et al. (1955). As a consequence of a confounding of variables the basis of Diven's conclusions were not clear and, as Chatterjee and Eriksen (1960) have pointed out, the findings of Lacey et al. could have been an artifact resulting from the method of computing the conditioning scores. The present results suggest that the conclusions of the Diven and Lacey et al. studies with respect to "learning without awareness" were not affected by their methodological shortcomings.

This study showed that emotionality as measured by the GSR to words associated with shock increased after a delay. This finding agrees with the results of Diven (1937), disagrees with those of Haggard (1943) and is only partially consistent with those of Mednick (1957). Mednick, it will be recalled, did not find an increase after a delay in emotional reactions to the CS, but did report such an increase for stimuli meaningfully related to the CS.

In the present experiment the increase in emotionality following a delay differed among the responses to SA, SU, and NS words with the increase to the SA words being the greatest. This result is in contradiction to Diven's (1937) conclusion that the increase following a delay is greatest under SU conditions. On the assumption that "incubation" is a result of recovery from adaptation, it was hypothesized that any increases after a delay should be greater for those words whose GSR showed greater adaptation during acquisition. Adaptation from the second to the last trial of acquisition was found to be greater for the GSR to SA than SU words, though the difference was not highly significant ($p < .10$).

Since both adaptation and the increase after a delay were found to be greater for SA words, these findings can be considered as consistent with the interpretation of incubation as recovery from adaptation. Also, the finding that subjects who showed relatively greater adaptation also showed a greater increase after a delay further supports this interpretation.

[4] As Eriksen (1960) has pointed out, the definition of awareness is a major methodological problem in this area. A similar problem involving the concept "emotion" has been discussed by Brown and Farber (1951). They have noted that to regard emotion as an existential real only partially encompassed by its definition rather than as a scientific inference has lead to unanswerable questions such as, what is the "true" measure of emotion. Their comments might also be considered applicable to the definition of awareness.

If one assumes that incubation is a result of recovery from adaptation, the question is raised as to what are the variables involved in adaptation. Figure 1 shows that despite the differential response strength among SA, SU, and NS words, response strength to all three words declined with increasing reinforcement. The shape of these curves suggests that the incubation effect may be considered as a special case of the Switzer phenomenon described by Hull (1943). Hull hypothesized that the initial rise in the extinction curve following massed reinforcement resulted from dissipation of some kind of inhibitory process developed during acquisition (inhibition of reinforcement). Within this framework one could speculate as to what kinds of inhibitory processes might be involved in the adaptation of emotional responses.

SUMMARY

The purpose of the study was to test the view that conditioned emotionality, as measured by the GSR, increases after a delay and that this increase is greater when the subject is aware of the relation between the CS and UCS. Certain words were associated with shock. The subject was aware of the relation between one word and shock but was unaware of the relation between a second word and shock. The GSR to these words was compared to the GSR to a word not associated with shock. The GSR to these words was recorded immediately after acquisition for half the subjects, and after 30 minutes delay following acquisition for the remaining subjects. The major results were: (*a*) the GSR increased as a function of delay, (*b*) the increase after a delay was greater when the subject was aware of the relation between the CS and UCS, (*c*) adaptation during acquisition was greater when the subject was aware of the CS-UCS relation. The results were interpreted as consistent with the interpretation of incubation as a result of recovery from adaptation.

REFERENCES

BINDRA, D., & CAMERON, L. Changes in experimentally produced anxiety with passage of time: Incubation effect. *J. exp. Psychol.*, 1953, **45,** 197–203.

BITTERMAN, M. E., & HOLTZMAN, W. N. Conditioning and extinction of the galvanic skin response as a function of anxiety. *J. abnorm. soc. Psychol.*, 1952, **47,** 615–623.

BROWN, J. S., & FARBER, I. E. Emotions conceptualized as intervening variables with suggestions toward a theory of frustration. *Psychol. Bull.*, 1951, **48,** 465–495.

CHATTERJEE, B. B., & ERIKSEN, C. W. Conditioning and generalization of GSR as a function of awareness. *J. abnorm. soc. Psychol.*, 1960, **60,** 396–403.

DIVEN, K. E. Certain determinants in the conditioning of anxiety reactions. *J. Psychol.*, 1937, **3,** 291–308.

ERIKSEN, C. W. Discrimination and learning without awareness: A methodological survey and evaluation. *Psychol. Rev.*, 1960, **67,** 279–300.

HAGGARD, E. A. Experimental studies in affective processes: I. Some effects of cognitive structure and active participation on certain autonomic reactions during and following experimentally induced stress. *J. exp. Psychol.*, 1943, **33,** 257–284.

HILLS, BARBARA. The role of awareness in verbal conditioning. Unpublished master's thesis, State University of Iowa, 1959.

HULL, C. L. *Principles of behavior*. New York: Appleton-Century-Crofts, 1943.

HUNTER, T. A., & Brown, J. S. A decade-type electronic interval-timer. *Amer. J. Psychol.*, 1949, **62,** 570–575.

LACEY, J. L., SMITH, R. C., & GREEN, A. Use of conditioned autonomic responses in the study of anxiety. *Psychosom. Med.*, 1955, **17,** 208–217.

MEDNICK, MARTHA T. Mediated generalization and the incubation effect as a function of manifest anxiety. *J. abnorm. soc. Psychol.*, 1957, **55,** 315–321.

MOELLER, G. The CS-UCS interval in GSR conditioning. *J. exp. Psychol.*, 1954, **48,** 162–166.

WHITE, R. W. *The abnormal personality*. New York: Ronald, 1948.

(Received November 18, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 63, No. 3, 540-545

REGRESSION AND DISORDERS OF THOUGHT[1]

LOREN J. CHAPMAN
University of Kentucky
ALVIN G. BURSTEIN
University of Michigan
DOROTHY DAY AND PAUL VERDONE
University of Chicago

MANY well known clinical writers have suggested that the disorders of thought displayed by schizophrenic or brain damaged patients represent in some sense a regression to modes of thought that are typical of young children. This statement has been made of schizophrenics by Arieti (1955), Kasanin (1945), Sullivan (1924), Von Domarus (1944), and White (1928); of brain damaged patients by Levin (1930); and of both schizophrenic and brain damaged patients by Hanfmann (1953), Lewin (1951), Vigotsky (1934), and Werner (1940). These writers were all concerned with the return to earlier patterns of interpreting or interrelating words and ideas, the sense of regression that is pertinent to this paper.

In psychoanalytic terminology, this is a variety of temporal ego regression. By ego regression, Freud (1953, 1955) referred to an abandonment by the ego of some of its functions. Although he mentioned that ego regression also entails a reversion to genetically earlier patterns of functioning, his examples of this were almost exclusively of the transformation of thoughts into images. When Freud wrote of temporal regression, he was usually concerned with its affective manifestations, i.e., libido regression and object regression. However, later psychoanalytic writers, like Kris (1936, 1938), have discussed in some detail temporal regression in the sense of the present paper, i.e., reversion to prelogical "primary process" patterns of interpreting and interrelating meaningful content.

In Lewin's (1951) terms, the phenomena under discussion are regarded as regression rather than retrogression because there is no evidence that any particular patient who shows these modes of thought previously exhibited them in his own childhood.

Attempts at objective testing of the regression hypothesis of pathological thinking have usually consisted of comparisons of the performance of children with that of psychiatric patients. The findings have been conflicting. Using several tests, Vigotsky (1934) concluded that children, schizophrenics, and brain damaged patients share a similar preconceptual kind of thinking. Bolles (1937) and Wegrocki (1940) each compared schizophrenics with children on a variety of tasks designed to measure the ability to think conceptually. Both investigators, contrary to Vigotsky, concluded from a qualitative analysis of performance that the differences between groups were much more prominent than the commonalities. Rashkis, Cushman, and Landis (1946) did a similar study in which they compared both schizophrenics and paretics with children. They concluded that the paretics but not the schizophrenics were childlike in their thinking.

All four of these studies relied on qualitative analysis of the data in terms of descriptive categories of response which were subjectively determined after the data had been obtained. This methodologically hazardous procedure often results in findings that fail to be confirmed by other investigators.

Several other investigators have employed more objective measures of performance. Cameron (1938), comparing schizophrenics with children on the ways in which they finish incomplete causal sentences, concluded that the groups are very different. However, three other investigators, using different tasks, reported similarities between these two groups. Feifel (1949) found that the definitions of words offered by schizophrenics and children showed a comparable pattern of departure from the definitions offered by adult normals.

[1] This study was supported by a research grant (M-3481) from the National Institute of Mental Health, United States Public Health Service.

Ellsworth (1951) reported a similar finding on the frequency of usage of different parts of speech. Burstein (1959, 1961) found that both schizophrenics and children treat antonyms as if they were synonyms more than do normals.

Each of these three investigators chose his measure of deviation on the basis of its distinguishing either schizophrenics or children from normal adults. In each case, the tasks were such that both children and schizophrenics were characterized by deviant performance relative to normal adults. However, as a result of restrictions either on the response alternatives available or on the investigator's categorization of the responses, each task provided relatively few avenues along which deviations could occur. The tasks tended to measure undifferentiated deviation rather than different kinds of specific deviation. Therefore, the likelihood that schizophrenics and children would appear similar was very great.

A much more stringent test of the regression hypothesis entails the use of tasks that measure errors that are sufficiently specific to distinguish pathological groups. For example, one might choose measures on the basis of their ability to distinguish the thinking disorder of schizophrenia from that of brain damage, as well as their ability to distinguish pathological from normal groups. One could then compare the performance of the clinical groups on several such measures with the performance of children with more assurance that sufficiently distinct and varied opportunities for deviation have been provided.

Such an approach has an additional advantage. Despite the fact that differences between the thinking of schizophrenic and brain damaged patients appear well established on the basis of both everyday clinical observation and research evidence, the disorders of both groups have been claimed to represent a regression to childhood patterns of thought. This claim cannot be valid for both groups unless schizophrenics and brain damaged patients regress to different age levels that differ on error preferences. It is especially appropriate, therefore, to apply a stringent test of the regression hypothesis to these two particular pathological groups.

The present investigation employs two tasks, each of which measures two contrasting kinds of errors. Each task has previously been shown to distinguish, by kind of error preferred, the thinking disorders of schizophrenic and brain damaged patients. One is a measure of the confusion of figurative and literal meanings of common words and phrases (Chapman, 1960); the second is a measure of the tendencies to include or exclude inappropriately the names of objects from common conceptual categories (Chapman, 1961).

These tasks present subjects with prearranged opportunities to engage in the hypothesized kinds of thinking disorder. It may be objected that this is inappropriate for investigating the hypothesis, since psychoanalysts write as if every occasion of regressive (primary process) thinking is motivated by affect specifically related to the material at hand. This emphasis may be in large part due to the fact that Freud's discussion of primary process thinking in adults arose primarily from his study of dreams, the psychology of errors, and neurosis. In psychosis, however, the thinking disorder apparently becomes a more pervasive feature of thinking. Burstein (1961) has shown that some errors structurally identical to those described as primary process thinking are exhibited by schizophrenics in dealing with experimentally contrived material which, presumably, is emotionally neutral. Such a process appeared in the tendency to treat both homonyms and antonyms as synonyms. These findings do not imply that the thinking disorder is not motivated, but they do suggest that some thinking disorders of the sort described by psychoanalytic writers are sufficiently pervasive to permit measurement through the use of materials that do not necessarily involve an individualistic investment of affect by the patient at hand.

EXPERIMENT I

Procedure

The confusion of figurative and literal meanings of common phrases was measured by a paper-and-pencil task with 40 items like the following:

30. The rich woman looks down on her servants.
This means:
A. The rich woman likes to work in her garden in the evening.
B. The rich woman sees her servants from a great height.

C. The rich woman thinks she is better than her servants.

16. The roof repairman looks down on the children playing.

This means:

A. The roof repairman likes to work in his garden in the evening.

B. The roof repairman sees the children from a great height.

C. The roof repairman thinks he is better than the children.

In Item 30 above, the correct answer is, of course, "The rich woman thinks she is better than her servants." Choosing "The rich woman sees her servants from a great height" would reflect an inappropriately literal misinterpretation of "looks down on." In Item 16, the reverse is true: "The roof repairman sees the children from a great height" is the correct answer, and choosing "The roof repairman thinks he is better than the children" would be inappropriately figurative. The third alternative, which interprets "looks down on" to mean to work in a garden, is neither correct nor a predicted error due to misinterpretation. This alternative was included as a measure of errors from such unidentified sources as inattention or inadequate understanding of the task. Such unidentified errors ought to be equally distributed among the three alternatives. Thus, there should be, by chance, as many errors made for unidentified reasons hidden among the predicted misinterpretation markings as there are markings of these "unidentified" error alternatives themselves. Therefore, we may obtain for each subject a measure of literal misinterpretation, relatively uncontaminated by errors from unidentified sources, if we subtract from the number of his literal error markings the number of his "unidentified" error markings. This will be called the "corrected literal misinterpretation score." The "corrected figurative misinterpretation score" is obtained by a similar principle.

The 40 items consisted of 20 which permitted literal misinterpretations and 20 which permitted figurative misinterpretations. For a more complete description of this task and its rationale, see Chapman (1960).

In an earlier study, schizophrenics were found to make literal misinterpretations of figurative usages more often than they made figurative misinterpretations of literal usages ($p < .01$). Organics, however, showed the opposite error perference, more often making figurative misinterpretations of literal usages ($p < .05$). Moreover, a preference score of figurative misinterpretations minus literal misinterpretations distinguished the two groups ($p < .01$).

Subjects

Subjects were pupils in a fourth grade class and an eighth grade class at a suburban public school near Chicago, Illinois. The students in this school were predominantly from upper middle class white families. There were 23 fourth graders, of whom 19 were 9 years old at their last birthway; 4 were 10 years old. There were 21 eighth graders, of whom 16 were 13 years old and 5 were 14.

TABLE 1

MEAN FIGURATIVE AND LITERAL MISINTERPRETATION SCORES OF FOURTH GRADERS AND EIGHTH GRADERS

	Figurative misinterpretations		Literal misinterpretations	
	Fourth	Eighth	Fourth	Eighth
Predicted misinterpretation markings	2.1	2.0	5.3	1.5
Unidentified error markings	1.3	.2	1.6	.3
Corrected misinterpretation score	.8	1.8	3.7	1.2

Results

The mean corrected literal misinterpretation scores and corrected figurative misinterpretation scores of the two groups are shown in Table 1. As seen there, the fourth graders made literal misinterpretations of figurative usages more often than they made figurative misinterpretations of literal usages. The difference was significant ($p < .01$) as determined by a two-tailed Wilcoxon test. The eighth graders tended to show the opposite error preference, preferring figurative misinterpretations more than literal misinterpretations, but this difference was not significant ($p = .19$). On a preference score of corrected figurative misinterpretation score minus corrected literal misinterpretation score, the two groups were again different ($p < .02$). Thus, the schizophrenics appeared to resemble the younger children in their performance on this test, whereas the brain damaged patients clearly did not.

EXPERIMENT II

Procedure

The subjects described in Experiment I were also given, in paper-and-pencil form, a task that had previously been employed to compare schizophrenics with brain damaged patients on the breadth of the concepts that they use.

The task consisted of 12 subtests like the following: 30 words were listed, of which 10 were names of trees, 10 names of flowering plants, and 10 names of musical instruments. The printed instruction read, "Circle each word that names a *tree*. Cross out each word that does *not* name a tree."

The 12 subtests varied in number of words to be categorized and in the proportion of words that belonged in the named category. Altogether the 12 sub-

tests required the categorization of 480 names of objects of which 240 belonged in a named category and 240 did not. Thus, there were as many opportunities for incorrect inclusions in a named category as for incorrect exclusions. For a further description of this task, see Chapman (1961).

In an earlier study, schizophrenics and brain damaged patients were presented this task in a card sorting format. The brain damaged patients made more exclusion errors than inclusion errors ($p < .05$), and the schizophrenics showed the opposite error preference, making more inclusion than exclusion errors ($p < .05$). Moreover, a preference score of inclusion errors minus exclusion errors distinguished the two groups ($p < .01$).

In the present study, this task was given to the fourth grade and eighth grade classes described above. It was completed by 20 subjects in each group.

Results

The mean exclusion and inclusion error scores are shown in Table 2. Both groups excluded more than they included. A score of inclusion errors minus exclusion errors was computed for each subject. Because random errors should result as often in an inclusion as an exclusion error, this difference score is independent of such random error. The difference score was found to distinguish the two groups ($p < .01$), the younger children showing the more extreme negative scores. Moreover, the younger children tended to exceed the older on the proportion of total errors that were exclusion errors ($p < .06$). (Subjects who made fewer than a total of five errors were omitted from this analysis because of the unreliability of proportions with small denominators.) Thus, the brain damaged patients appeared like the children in their performance, but more like the younger children than the older. The schizophrenics exhibited the opposite error preference on this test. The results here are essentially contrary to those obtained with the figurative-literal test.

TABLE 2
MEAN ERRORS ON A MEASURE OF CONCEPTUAL BREADTH

	Inclusion errors	Exclusion errors	Inclusion minus Exclusion
Fourth grade	8.6	37.0	−28.4
Eighth grade	4.6	8.6	−4.0

DISCUSSION

On two tasks that distinguish schizophrenics from brain damaged patients, younger children were more like schizophrenics in their error preference on one task; on the second, they were more like brain damaged patients. These results appear contradictory if one accepts the regression hypothesis for either or both clinical groups. This contradiction cannot be resolved by hypothesizing different age levels of regression for schizophrenics and brain damaged patients because each group of patients would appear regressed to an earlier level than the other in terms of one of the two tasks. Perhaps, then we should not look to the regression principle to explain the similarity of either clinical group to the younger children.

One may speculate that the error preferences which occur as a function of age with the children on these tasks result from the smaller fund of knowledge possessed by younger children. This interpretation might hold for the figurative-literal task. Figurative meanings may be learned later in life than literal ones, so younger children misinterpret figurative expressions as if they were literal more often than they make the opposite kind of misinterpretation. On the breadth of concepts task, younger children may more often exclude objects from the named conceptual categories because they have not yet learned the names of the objects. But this explanation in terms of childish ignorance would not explain the results for the clinical groups because the results are in opposite directions for the two tasks. The data offer no justifiable basis for speculation as to the theoretical basis for the differing error patterns exhibited by the pathological groups. They do, however, demonstrate the inadequacy of the hypothesis that thinking disorders are completely accounted for by regression.

As previously pointed out, an investigator is more likely to conclude that the errors of two groups are qualitatively alike if his measure is not highly specific as to the kind of error it taps. One may enhance the specificity of description of thinking disorder if he has available two or more measures on the same subjects and treats them concurrently and

interrelates them. Using this procedure, one is less likely to conclude that groups are similar than if the measures are examined one at a time. This is because many of the same kinds of deviant responses and failures appear in a variety of conditions. The difference between thinking disorders in diverse conditions partly consists of differing frequencies of various kinds of deviant response which are present in many of the conditions. For example, on the second task here, schizophrenics, unlike the younger children, excluded less than they included. However, considering exclusion errors alone, one finds that the schizophrenics made more exclusion errors than the normal adults, and the younger children made more than the older children. The difference between the groups emerged only in rank ordering the strengths of two *differing* error propensities.

The use of rank ordering of different kinds of response in order to specify a difference between children and schizophrenics is also the principal evidence in Cameron's (1938) well known study. He compared three groups of schizophrenics, varying in degree of pathology on their handling of two tasks which, according to Piaget, require abilities that mature at different ages. The more disorganized patients did not do worse on the tasks that required an ability which matures later. Cameron concluded, therefore, that the schizophrenic thinking disorder is not regressive in nature. However, if Cameron had only been concerned with the presence or absence of the various kinds of deviant response, he might well have concluded that children and schizophrenics are alike. Indeed, he remarked that both show asyndesis and metonymic distortion.

Even in the four rather subjectively interpreted studies of concept sorting that were mentioned earlier, the degree to which the investigator particularized his description of the thinking disorder seemed to influence his conclusions concerning regression. Vigotsky (1934) apparently lumped together into the single category of "complex thinking" a fairly large variety of failure on his test and concluded that his groups were alike. Bolles (1937), Wegrocki (1940), and Rashkis, Cushman, and Landis (1946) described sortings in terms of more varied categories, concluding that the differences between groups were more important than the similarities.

The basic question would seem to be whether regression is to be considered a loose descriptive term or an explanatory principle. One may justifiably use it descriptively as a way of referring to whatever correspondences emerge between two sets of observations, one of which is obtained on children. Very often, however, the term is used as an explanatory principle, meaning, for example, that patients behave as they do because they have become like children. This usage seems to require a fairly blanket similarity between the two groups unless one is willing to say that regression occurs in certain characteristics but not others. Such a formulation is rather unsatisfactory in the absence of a rationale for predicting which childlike characteristics are shown.

SUMMARY

The hypothesis that the thinking disorders of schizophrenics and brain damaged patients are forms of regression was tested by administering to a fourth grade and an eighth grade class of children two tasks, each of which had previously been shown to distinguish schizophrenics from brain damaged patients on the variety of error most often made. The first task measured both the misinterpretation of literal usages of words as if they were figurative, and the misinterpretation of figurative usages as if they were literal. The younger children exceeded the older in their differential preference for misinterpreting figurative expressions as if they were literal as against the opposite kind of misinterpretation. Since schizophrenics had previously been shown to differ from brain damaged patients in the same manner, the finding indicated that schizophrenics are more "childlike" than brain damaged patients on this task.

The second task measured inappropriate inclusion and exclusion of objects from common conceptual categories. The younger children exceeded the older on their preference for exclusion errors as against inclusion errors. Brain damaged patients had previously been shown to differ from schizophrenics in the same manner, indicating that the brain

damaged are more "childlike" than the schizophrenics on this task.

There is no blanket similarity between the error patterns of children and of either schizophrenics or brain damaged patients, and the use of the term "regression" to imply such a blanket similarity is not justified. There are, however, some points of similarity, which would appear to justify the use of regression as a purely descriptive term, as distinct from an explanatory concept. Much of the lack of agreement among investigators in this area has occurred both because the various investigators have differed as to how much similarity is necessary to conclude that regression exists, and because they have used instruments that differ as to sensitivity for distinguishing varieties of deviancy.

REFERENCES

ARIETI, S. *Interpretation of schizophrenia.* New York: Brunner, 1955.

BOLLES, MARJORIE M. The basis of pertinence. *Arch. Psychol., N. Y.*, 1937, **30**, No. 212.

BURSTEIN, A. G. Primary process in children as a function of age. *J. abnorm. soc. Psychol.*, 1959, **59**, 284–286.

BURSTEIN, A. G. Some verbal aspects of primary process thought in schizophrenia. *J. abnorm. soc. Psychol.*, 1961, **62**, 155–157.

CAMERON, N. Reasoning, regression, and communication in schizophrenics. *Psychol. Monogr.*, 1938, **50** (1, Whole No. 221).

CHAPMAN, L. J. Confusion of figurative and literal usages of words by schizophrenics and brain damaged patients. *J. abnorm. soc. Psychol.*, 1960, **60**, 412–416.

CHAPMAN, L. J. A reinterpretation of some pathological disturbances in conceptual breadth. *J. abnorm. soc. Psychol.*, 1961, **62**, 514–519.

ELLSWORTH, R. B. The regression of schizophrenic language. *J. consult. Psychol.*, 1951, **15**, 387–391.

FEIFEL, H. Qualitative differences in the vocabulary response of normals and abnormals. *Genet. psychol. Monogr.*, 1949, **39**, 151–204.

FREUD, S. Metapsychological supplement to the theory of dreams. (Originally published 1916) In, *Collected papers.* Vol. 4. London: Hogarth, 1953. Pp. 137–151.

FREUD, S. *The interpretation of dreams.* (Originally published 1900) New York: Basic Books, 1955.

HANFMANN, EUGENIA. Concept formation test. In A. Weider (Ed.), *Contributions toward medical psychology.* Vol. 2. New York: Ronald, 1953. Pp. 731–740.

KASANIN, J. S. Developmental roots of schizophrenia. *Amer. J. Psychiat.*, 1945, **101**, 770–776.

KRIS, E. The psychology of caricature. *Int. J. Psycho-Anal.*, 1936, **17**, 285–303.

KRIS, E. Ego development and the comic. *Int. J. Psycho-Anal.*, 1938, **19**, 77–90.

LEVIN, M. Archaic regressive phenomena as a defense mechanism in schizophrenia. *Arch. Neurol. Psychiat.*, 1930, **24**, 950–965.

LEWIN, K. *Field theory in social science.* New York: Harper, 1951.

RASHKIS, H. A., CUSHMAN, JANE F., & LANDIS, C. A new method for studying disorders of conceptual thinking. *J. abnorm. soc. Psychol.*, 1946, **41**, 70–74.

SULLIVAN, H. S. Schizophrenia: Its conservative and malignant features. *Amer. J. Psychiat.*, 1924, **81**, 77–91.

VIGOTSKY, I. S. Thought in schizophrenia. *Arch. Neurol. Psychiat.*, 1934, **31**, 1063–1077.

VON DOMARUS, E. The specific laws of logic in schizophrenia. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia.* Berkeley: Univer. California Press, 1944. Pp. 703–714.

WEGROCKI, H. J. Generalizing ability in schizophrenia: An inquiry into the disorders of problem thinking in schizophrenia. *Arch. Psychol., N. Y.*, 1940, **36**, No. 254.

WERNER, H. *Comparative psychology of mental development.* New York: Harper, 1940.

WHITE, W. A. The language of schizophrenia. *Res. Publ. Ass. Res. Nerv. Ment. Dis.*, 1928, **5**, 323–343.

(Received November 21, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 63, No. 3, 546-551

EFFECTS OF MEPROBAMATE AND OF PROCHLORPERAZINE ON POSITIVE AND NEGATIVE CONDITIONING

LEONARD UHR, MARGARET CLAY, ARTHUR PLATZ, JAMES G. MILLER, AND E. LOWELL KELLY

University of Michigan

THE use of conditioning techniques in the study of the psychological effects of centrally acting drugs is important from three points of view. First, for the learning theorist, drive states may be manipulated experimentally by drugs and thus may be a useful tool for testing deductions about the effects of drive level on the rate of learning (Brown, DiMascio, & Klerman, 1958; Burnstein & Dorfman, 1959). Second, for the physiologist, the effects of drugs on different kinds of conditioning make possible inferences about the site of drug action (Berger, 1957; Hunt, 1957) and may provide data on the manner in which different drugs affect specific brain mechanisms (Olds & Travis, 1960). Third, conditioning techniques provide useful information for the psychopharmacologist and the clinician on how different drugs alter drives (Schneider & Costiloe, 1957).

Previous studies have generally found that depressant drugs retard the acquisition of simple conditioned responses and speed extinction, while stimulants facilitate acquisition and slow extinction. (For a brief summary of previous work, see Schneider & Costiloe, 1957.) The present study attempts to demonstrate that the drug effect is partially determined by whether the affective state aroused by the unconditioned stimulus (UCS) used is positive or negative. In the basic Hullian paradigm, the tendency to respond (E or excitatory reaction potential) is a function of the drive level (D) times habit strength (H, or amount of learning) minus the response threshold (L): $E = (D \times H) - L$. Since the number of trials to acquisition or extinction is a response measure and not a direct measure of the amount of learning, different theoretical interpretations of the results are possible. Thus Eysenck (1960), for instance, seems to feel that drugs act directly on habit strength (H). Payne (1958) assumes that depressant drugs increase the excitation threshold of neural structures and thus raise the reaction threshold (L). The most common interpretation of these experimental results, however, assumes that the reaction potential is affected because drugs act directly on the drive level (D) of the organism. The experimental designs of most of these studies (like the present one) do not provide a crucial test between these alternative interpretations. But since there is abundant clinical evidence that tranquilizing drugs do decrease hyperexcitability and anxiety (Berger, Hendley, & Tynes, 1956; Osinski, 1957; Wikler, 1958) and some experimental data showing they also decrease the activation level as measured by the psychogalvanometer (Carran, 1960; Kristofferson & Cormack, 1960), our predictions are based on the assumption that tranquilizing drugs possess drive-altering properties.

This series of two experiments was designed to test the effects of two tranquilizers (meprobamate and prochlorperazine) on the conditioning of the galvanic skin response (GSR) when a noxious stimulus is used as the UCS (shock in Experiment I) and when an affectively positive stimulus is used as the UCS (sexually exciting pin-up pictures in Experiment II). The GSR was chosen as the unconditioned response (UCR) because it provides a sensitive index of the "activation level" of the organism (Woodworth & Schlosberg, 1954) and can provide information on the effects of tranquilizers upon autonomic activity. The procedure used was adapted from Schneider and Costiloe (1957) with slight modifications.

EXPERIMENT I: SHOCK CONDITIONING

Subjects

The subjects were college men, 21 years of age and older, who volunteered for the experiment and who were paid for their participation. In Experiment I, 27 subjects were tested, 9 each under meprobamate, prochlorperazine, and placebo.

Dosage

On the second experimental day, one hour before testing began, subjects were administered capsules containing either 1200 mg. meprobamate, 20 mg. prochlor-

perazine, or a placebo. All subjects received identical capsules; administration followed double-blind procedures.

Apparatus

The GSR was recorded from the left palm on a Lafayette psychogalvanometer. The UCS was a shock administered to the right arm by a standard ac shocker. Shock intensity varied from 50 to 70 volts between subjects according to individual tolerance, but was maintained at a constant level for each subject on both days of testing. The pulse frequency was 120 cycles per second. On the oscilloscope, the pulse was sharply peaked with a fast fall-off. The shocker did not have the ability to keep current constant over side changes in the subject's resistance. The subjects were seated in a modified dental chair and completely screened from the experimenter. In order to cut down on distractions, the experimental room was kept at low illumination during testing.

Procedure

Following Schneider and Costiloe (1957), a partial reinforcement, short-trace conditioning procedure was used. The conditioned stimulus (CS) was a buzzer of 1 second duration, followed in 0.5 second by an electric shock. The duration of the UCS was 2 seconds. The presentation of the stimuli was controlled by an electronic timer.

Each subject was tested on two consecutive days, without drugs the first day and with drugs the second. On each day, a four-step procedure was followed:

First Day

1. The GSR to the CS was adapted out. The criterion used for adaptation and extinction was three consecutive trials without response to the CS or until the nonreinforced CS had been presented for 30 trials.

2. The UCS was presented for five trials at 30-second intervals to test the subject's reaction.

3. The partial reinforcement conditioning procedure was begun. Presentation of the CS occurred every 30 seconds for 20 trials. On 15 of the trials, the CS was followed by the UCS, and on 5 it was not. The nonreinforced trials were randomized according to the same schedule for all subjects within the series of 20 trials, and changed for each subject on the second day of testing.

4. Then the extinction procedure was carried out. The CS was again presented every 30 seconds, but without reinforcement, until extinction or for 30 trials. The number of trials to extinction (E1) provides a rough measure of the strength of conditioning and, when possible, was used in matching subjects assigned to the different experimental groups on the second day.

Second Day

5. The CS was again presented without reinforcement as in Step 4, until extinction occurred. The second extinction (E2) provides the measure of the amount of spontaneous recovery under each drug condition.

6. Presentation of UCS for five trials as in Step 2.

7. Reconditioning as in Step 3.

8. Extinction as in Step 2. The third extinction (E3) provides an index of the amount of reconditioning on the second day.

In order to control for individual differences in the amount of original conditioning, a measure of spontaneous recovery was obtained for each subject by subtracting the number of trials for E2 from the number for E1 (E1 − E2). Subtracting the number of trials for E3 from E1 provided the measure of the amount of reconditioning (E1 − E3). The statistical analysis compares the means of these difference scores for the two drug groups with the difference scores for the control group.

Predictions

Although the two drugs tested are both ataraxics, their mode of operation and site of action are completely different. Prochlorperazine belongs to the group of tranquilizers called the "autonomic suppressants," which affect the internal regulating systems by stimulating the posterior hypothalamus. Meprobamate, in contrast, is a central relaxant and produces its tranquilizing effect by a selective blocking of interneuronal circuits, resulting in a reduction in muscle tension (Berger, 1957).

Different behavioral effects of these drugs parallel these neurophysiological differences. Sikorski (1958) found that both meprobamate and prochlorperazine facilitate the expression of sexual and aggressive impulses, apparently by reducing the anxiety generally associated with their expression. Olds and Travis (1960), however, found quite striking differences between the effects of meprobamate and chlorpromazine (a phenothiazine) on self-stimulation behavior. While the former inhibited escape and avoidance reactions in doses that did not affect approach tendencies, the latter inhibited both approach and avoidance responses such that self-stimulation was markedly reduced.

The predictions for the shock conditioning are based on the assumption that both meprobamate and prochlorperazine relieve anxiety and thus lower the over-all drive level of the subject. Since reaction potential is a multiplicative function of habit strength and drive ($E = D \times H$), and since the habit strengths of the drug and placebo groups are assumed equal because of the identical criterion for extinction on the first day, the two drug groups should show less spontaneous recovery

than the comparable control group. The differences, however, should be slight because the habit strength of a previously extinguished response is small.

The same reasoning leads to the prediction that the drug groups should extinguish faster than the control group after relearning, and that the difference should be larger than for the extinction of the spontaneous recovery because of greater habit strength after conditioning than after extinction.

Results

Table 1 presents the mean number of trials needed for each extinction as well as the measures for the amount of spontaneous recovery (E1 − E2) and the amount of reconditioning (E1 − E3). Table 2 presents the statistical tests of significance for these differences.

On the first day, when no drug was administered, the three groups showed no significant difference in the number of trials needed to reach extinction. It was hypothesized that the administration of meprobamate and prochlorperazine on the second day would result in faster extinction of the spontaneous recovery in the two drug groups as compared to the placebo group. The meprobamate group was in the predicted direction; the prochlorperazine group was not. Neither difference was significant. The prediction that the drug group would show faster extinction at the end of the second testing also was not confirmed, although both drug groups were in the predicted direction. Experiment I, then, gives no evidence that tranquilizing drugs affect classical conditioning or spontaneous recovery when a noxious UCS is used as the reinforcer.

TABLE 1
MEAN NUMBER OF TRIALS TO EXTINCTION UNDER SHOCK CONDITIONING

Group	First extinction (E1)	Spontaneous recovery (E2)	E1 − E2	Final extinction (E3)	E1 − E3
Placebo	18.3	15.0	3.3	12.2	6.1
Meprobamate	17.6	13.0	4.6	9.3	8.3
Prochlorperazine	21.9	20.2	1.7	12.9	9.0

TABLE 2
TESTS OF DRUG EFFECTS ON SHOCK CONDITIONING

Baseline	Response measure	t^a
Meprobamate vs. Placebo groups (First day: without drugs)	Amount of conditioning (E1)	.11
Prochlorperazine vs. Placebo groups (First day: without drugs)	Amount of conditioning (E1)	.65
Drug Group		
Meprobamate vs. Placebo	Spontaneous recovery (E1 − E2)	.30
Meprobamate vs. Placebo	Reconditioning (E1 − E2)	.46
Prochlorperazine vs. Placebo	Spontaneous recovery (E1 − E2)	.33
Prochlorperazine vs. Placebo	Reconditioning (E1 − E3)	.59

[a] With 16 *df* a *t* of 2.12 is needed for significance at the 5% level.

EXPERIMENT II: PIN-UP PICTURE CONDITIONING

Since Olds (Olds & Travis, 1960) has presented evidence that drugs may affect the reward areas of the brain differently from the regions mediating escape tendencies, it was decided to run a second experiment, using an affectively positive stimulus as the UCS. On the basis of their work with self-stimulation techniques, Olds and Travis concluded that dosages of meprobamate sufficiently large to inhibit escape and avoidance tendencies had little or no effect on approach tendencies. We tried to capitalize on a "built-in" approach-avoidance conflict situation by using sexually arousing pictures as the unconditioned stimuli for the affective arousal measured by the GSR. It was hypothesized that meprobamate would lower the avoidance aspect of the conflict associated with the arousal of sexual impulses in social situations. This would allow greater expression of sexual affect (and thus greater drive level) and, therefore, greater spontaneous recovery and higher reaction potential built up during relearning. Two experiments lend support to the contention that the degree of sexual arousal may be accentuated by lessening inhibition. Clark (1952) found that sexual imagery in responses made to TAT cards was both more frequent and more direct after subjects had taken alcohol. Sikorski (1958) similarly found that both need for aggression and need for sex were higher, as measured by objectively scored responses to TAT cards, for subjects under

either meprobamate or prochlorperazine than for a comparable placebo group.

It was felt that the sex of the experimenter would be an important determinant of the amount of sexual arousal, and it was predicted that under meprobamate, male subjects run by a female experimenter would show slower extinction than male subjects run by a male experimenter.

Subjects

Fifty subjects were used in all. The female experimenter ran 10 subjects in each of the three groups: meprobamate, prochlorperazine, and placebo. The male experimenter ran 10 subjects in both the meprobamate and placebo groups.

Apparatus

The apparatus was that used in Experiment I with the exception that the UCS were mounted pin-up pictures which were viewed at the end of a large box through a 4 × 5-inch opening. The pictures were mounted approximately 17 inches from the aperture. A small light was placed on the roof of the box near the front end. The UCS was illuminated by a flash of light lasting 2 seconds. At the end of each presentation, the picture was changed for the next exposure. The experimental room was kept dark during the conditioning procedure so that subjects were unable to see the picture until the light was flashed.

Procedure

A 2-day testing procedure identical to that described for Experiment I was used.

Results

Table 3 presents the mean number of extinction trials required for each of the three extinction periods. The significance of these differences is presented in Tables 3, 4, and 5.

The prochlorperazine group as compared to the control group showed faster extinction of the spontaneous recovery and slower extinction of the second day's relearning, but neither difference approached significance. In the comparison of the meprobamate groups with the placebo groups, the amount of conditioning the first day, as measured by the number of trials to extinction, was not significantly related to either the sex of the experimenter or the subsequently assigned drug group. The difference between E1 and E2 is smaller for the total meprobamate group than for the control group, and smaller for the subjects run by the female experimenter than for those run by the male experimenter. However, these predicted differences are not significant. The E1 − E3 difference is significant in the predicted direction at the .05 level for both drug and experimenter effects. There is no drug-experimenter interaction. The meprobamate group shows comparatively slower extinction at the end of the second day, and those drug subjects run by the female experimenter were extinguished more slowly than those run by the male experimenter.

TABLE 3

MEAN NUMBER OF TRIALS TO EXTINCTION UNDER SEXUAL AROUSAL CONDITIONING

Group	E1	E2	E1 − E2	E3	E1 − E3
Meprobamate					
Male experimenter	12.9	9.2	3.7	8.3	4.6
Female experimenter	5.6	6.0	−.4	10.8	−5.2
Placebo					
Male experimenter	14.8	16.4	−1.6	4.3	10.5
Female experimenter	15.3	9.1	6.2	9.5	5.8
Prochlorperazine					
Male experimenter	—	—	—	—	—
Female experimenter	16.4	8.6	7.8	12.8	3.6

TABLE 4

ANALYSIS OF VARIANCE OF MEPROBAMATE EFFECTS UNDER SEXUAL AROUSAL CONDITIONING

Baseline	Response measure	*F*
Meprobamate vs. Placebo (First day: without drugs) Male vs. Female experimenter	Amount of conditioning (E1)	1. Drug group effect: 3.27 2. Experimenter: 1.13 3. Interaction: 1.48
Drug effect A. Meprobamate vs. Placebo Male vs. Female experimenter	Spontaneous recovery (E1 − E2)	1. Drug effect: .02 2. Experimenter: .19 3. Interaction: 2.00
B. Meprobamate vs. Placebo Male vs. Female experimenter	Reconditioning (E1 − E3)	1. Drug effect: 6.33* 2. Experimenter: 4.66* 3. Interaction: .58

* Significant beyond the 5% level, $df = 1/36$.

TABLE 5
t TESTS OF PROCHLORPERAZINE EFFECTS UNDER SEXUAL AROUSAL CONDITIONING

Drug group	Response measure	*t*[a]
Baseline		
Prochlorperazine vs. Placebo (First day: without drugs)	Amount of conditioning (E1)	.06
Drug effect		
Prochlorperazine vs. Placebo	Spontaneous recovery (E1 − E2)	.23
Prochlorperazine vs. Placebo	Reconditioning (E1 − E3)	.41

[a] $df = 18$. A t of 2.10 is needed for significance at the .05 level.

TABLE 6
COMPARISON OF SPEED OF EXTINCTION WITH POSITIVE AND NEGATIVE UNCONDITIONED STIMULI

Group	Main effects	*F*
Placebo[a]	Spontaneous recovery vs. Reconditioning	.15
	Shock vs. Pin-up pictures	.12
	Interaction	.22
Prochlorperazine[a]	Spontaneous recovery vs. Reconditioning	.07
	Shock vs. Pin-up pictures	.01
	Interaction	1.57
Meprobamate[a]	Spontaneous recovery vs. Reconditioning	.02
	Shock vs. Pin-up pictures	7.73*
	Interaction	1.59

[a] Includes only subjects run by the female experimenter.
* Significant beyond the .01 level, $df = 1/32$.

Since it was predicted that the drugs would facilitate extinction if an aversive drive were aroused but delay extinction if the aroused drive were affectively positive, the extinction rates under positive and negative conditioning were compared within each drug group (Table 6). The placebo groups showed no difference in extinction rate attributable to the UCS, nor did the prochlorperazine groups. The meprobamate groups, however, were significantly different beyond the .01 level, with the group conditioned to pin-up pictures taking longer to extinguish than the group conditioned with shock. The expected faster extinction of the spontaneous recovery did not occur, perhaps because the second day's procedure was identical to that used the first day and thus alerted or allowed the subjects to become set for the extinction procedure.

The data of Experiment II are consistent with the hypothesis that meprobamate allows greater expression of sexual impulses, possibly by reducing internal inhibiting responses such as fear. The two experiments taken together are consistent with Olds' (Olds & Travis, 1960) finding that doses of meprobamate which inhibit avoidance reactions will not affect approach reactions, and with Sikorski's (1958) finding that meprobamate facilitates the expression of sexual effect. The failure of the E1 − E2 difference to be statistically significant can perhaps be accounted for in terms of Hullian theory. In both experiments, the predicted differences in the extinction rates between the drug and placebo groups was predicted on the assumption that the drugs affected the drive level of the organism and thus indirectly affected the reaction potential by combining multiplicatively with habit strength. However, after the first extinction session, the remaining habit strength was low, and, therefore the manipulation of drive on the second day had little effect. After reconditioning the second day, habit strength was high, and different drive levels would, by multiplying with habit strength, produce a relatively greater difference in the reaction potential between the two groups than after the first day. Inspection of Tables 1 and 3 shows this to be indeed the case: In all instances there is a greater difference between the drug and placebo groups for E1 − E3 than for E1 − E2, as predicted, though the difference is not significant within drug groups taken separately.

SUMMARY

The present study determined the effects of two tranquilizing drugs (meprobamate and prochlorperazine) on the retention and relearning (by means of conditioning procedures) of an emotional response to noxious and positive stimuli. Seventy male subjects were tested twice on consecutive days. The first day the GSR was conditioned and extinguished. On the second day, one hour after double-blind administration of either a drug or placebo, the subjects were again extinguished, reconditioned, and extinguished again. Two experiments were run, one using as the UCS a negative stimulus (shock) and one using a positive stimulus (pin-up pictures).

With a negative UCS, neither drug affected at a statistically significant level the number of trials to extinction after spontaneous recovery, or the number required for reconditioning. With a positive UCS, (*a*) prochlorperazine affected neither spontaneous recovery nor reconditioning; (*b*) meprobamate did not affect spontaneous recovery; (*c*) the number of trials to extinction of the reconditioning was significantly higher under meprobamate than under placebo; (*d*) among subjects who had taken meprobamate, those run by a female experimenter took significantly more trials to extinguish after reconditioning than did those run by a male experimenter; (*e*) subjects under meprobamate but not under prochlorperazine took longer to extinguish a response conditioned with a positive UCS than with a negative UCS.

Of the four predictions for the prochlorperazine vs. placebo groups, only two were in the predicted direction, and none was significant. All six predictions for the meprobamate vs. placebo groups were in the predicted direction, but only two comparisons were statistically significant: drug effect and experimenter differences after reconditioning to sexually arousing pictures. Subjects under meprobamate extinguished faster than the control group when the UCS was noxious, and more slowly than the control group when a positive UCS was used. The data do not indicate that the extinction of the spontaneous recovery was faster than the extinction of the reconditioning. The meprobamate groups, but not the placebo or prochlorperazine groups, extinguished the conditioned response to shock significantly faster than the response conditioned to pin-up pictures.

REFERENCES

BERGER, F. M. The chemistry and mode of action of tranquilizing drugs. *Ann. N. Y. Acad. Sci.*, 1957, **67**, 685–699.

BERGER, F. M., HENDLEY, C. D., & TYNES, T. E. Effect of mephrobamate (Miltown) on animal behavior. *Fed. Proc.*, 1956, **15**, 400. (Abstract)

BROWN, J. W., DIMASCIO, A., & KLERMAN, G. L. Exploratory study on the effects of phrenotropic drugs on competitive paired-associate learning. *Psychol. Rep.*, 1958, **4**, 583–589.

BURNSTEIN, E., & DORFMAN, D. Some effects of meprobamate on human learning. *J. Psychol.*, 1959, **47**, 81–86.

CARRAN, A. Drug effects and activation level. In L. Uhr & J. G. Miller (Eds.), *Drugs and behavior*. New York: Wiley, 1960. Pp. 409–412.

CLARK, R. A. The projective measurement of experimentally induced levels of sexual motivation. *J. exp. Psychol.*, 1952, **44**, 391–399.

EYSENCK, H. J. Drug postulates, theoretical deductions, and methodological considerations. In L. Uhr & J. G. Miller (Eds.), *Drugs and behavior*. New York: Wiley, 1960. Pp. 352–359.

HUNT, H. F. Some effects of meprobamate on conditioned fear and emotional behavior. *Ann. N. Y. Acad. Sci.*, 1957, **67**, 712–722.

KRISTOFFERSON, A. B., & CORMACK, R. H. Preliminary results concerning the galvanic skin response as an indicator of drug effects. In L. Uhr & J. G. Miller (Eds.), *Drugs and behavior*. New York: Wiley, 1960. Pp. 405–408.

OLDS, J., & TRAVIS, R. P. Effects of chlorpromazine, meprobamate, pentobarbital and morphine on self-stimulation. In L. Uhr & J. G. Miller (Eds.), *Drugs and behavior*. New York: Wiley, 1960. Pp. 255–267.

OSINSKI, W. A. Treatment of anxiety states with meprobamate. *Ann. N. Y. Acad. Sci.*, 1957, **67**, 766–772.

PAYNE, R. B. An extension of Hullian theory to response decrements resulting from drugs. *J. exp. Psychol.*, 1958, **55**, 342–346.

SCHNEIDER, R. A., & COSTILOE, J. P. Effect of centrally acting drugs on conditioning in man: The inhibiting and facilitating effects of chlorpromazine, amobarbital, and methylphenidylacetate on the conditioned galvanic skin response. *Amer. J. med. Sci.*, 1957, **233**, 418–423.

SIKORSKI, J. The effects of meprobamate and prochlorperazine on need for aggression and need for sex. Unpublished bachelor of arts honors thesis, University of Michigan, 1958.

WIKLER, A. *The relation of psychiatry to pharmacology*. Baltimore: Williams & Wilkins, 1958.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology*. New York: Holt, 1954.

(Received November 21, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 63, No. 3, 552-561

THE RELATIONSHIP OF PERSISTENCE AT A TASK TO EXPECTATION OF SUCCESS AND ACHIEVEMENT RELATED MOTIVES[1]

N. T. FEATHER[2]
University of Michigan

THE present study investigates the relationship of persistence at a task both to its apparent difficulty and to the relative strength within an individual of the motives to achieve success and to avoid failure (Feather, 1960). The situation to be considered is one in which a subject comes to an objectively insoluble puzzle either believing it to be easy or very difficult. He works at this achievement task under test conditions and suffers repeated and consistent failure in his attempts to get the solution. Persistence is measured in terms of the total time or total trials which the subject works at the task before he turns to an alternative achievement activity. The former measure is sometimes referred to in the literature as temporal persistence; the latter is analogous to resistance to extinction.

The hypotheses of the present investigation are developed from the theory of achievement motivation (Atkinson, 1957, 1960) together with certain additional assumptions. This theory relates characteristics of motivated behavior to the interaction of relatively stable personality dispositions (motives) and more transient situational influences (expectations and incentive values). As such, this approach belongs to a class of "expectancy value" theories which all involve somewhat similar concepts (Feather, 1959a). Within this context, the present research attempts to clarify the relationship of persistence at a task to expectations and motives by providing answers to the following two questions:

1. How is persistence at a task affected by the initial subjective probability of success among subjects in whom the motive to achieve success is stronger than the motive to avoid failure ($M_s > M_{af}$)? Do such subjects tend to persist longer at the task when it initially appears easy to them than when it initially appears very difficult?

2. Conversely, how is persistence at a task affected by the initial subjective probability of success among subjects in whom the motive to avoid failure is stronger than the motive to achieve success ($M_{af} > M_s$)?

The present investigation goes beyond the earlier studies of French and Thomas (1958) and Atkinson and Litwin (1960) in its attempt to vary both motive and expectation simultaneously. Furthermore, it provides assumptions about the dynamics of changing motivation as expectations of success and failure change, and an explicit analysis of why the subject stops performing the task.

THEORETICAL ASSUMPTIONS

Consider a subject who is performing the initial achievement task. It is assumed that he will continue to persist at this task as long as total motivation to perform it is stronger than total motivation to perform the alternative available to him.

Total motivation to perform the initial achievement task is attributable to the following component motivations[3]: (*a*) achievement related motivation to perform the task, and (*b*) extrinsic motivation to perform the task. Similarly, since the alternative is also an achievement task, total motivation to perform it is attributable to analogous component motivations.

Achievement related motivation to perform a task refers to the *resultant* of motivation to

[1] This paper is based upon a doctoral dissertation submitted to the department of psychology at the University of Michigan.

The research was part of a project on Personality Dynamics financed by a grant from the Ford Foundation.

The author wishes to thank John W. Atkinson, dissertation chairman, for his advice and criticism; and J. David Birch, Dorwin Cartwright, Clyde C. Coombs, and John B. Lansing for their suggestions.

[2] Now at the University of New England, Armidale, New South Wales, Australia.

[3] It should be noted that a distinction is made between motivation, which refers to a particular act or response, and motive. Differences in strength of motive (a latent disposition of personality) are assumed to be only one factor influencing the strength of motivation, the other influences being level of expectation and incentive value.

achieve success at the task and motivation to avoid failure at the task. According to the theory of achievement motivation (Atkinson, 1957, 1960) these two motivations summate algebraically to give *positive* achievement related motivation (approach) for subjects in whom the motive to achieve success is stronger than the motive to avoid failure (i.e., when $M_s > M_{af}$), and *negative* achievement related motivation (avoidance) for subjects in whom $M_{af} > M_s$. Further, achievement related motivation whether positive or negative is at its maximum when subjective probability of success at the task is intermediate (i.e., when $P_s = .50$) and decreases monotonically as P_s becomes either smaller or larger than .50. If achievement related motivation were the only motivation elicited in the situation it is apparent that a subject with stronger motive to avoid failure should not even undertake performance of an achievement task. Instead he should avoid the task and choose activities which do not arouse anxiety about failure. In contrast, a subject with stronger motive to achieve success should show some positive interest in performing an achievement task.

The concept of extrinsic motivation to perform a task is introduced to account for the fact that the subject is in a social situation in which he has the role of a subject in an experiment and knows he is expected to make some attempt at the task. In any test situation there are certain extrinsic constraints that influence the subject to perform the task irrespective of the nature of his achievement-related motivation. Studies by French (1955), and by Atkinson and Raphelson (1956) have shown, for example, that other motives like n Affiliation are sometimes systematically related to task performance in a situation where no achievement orientation is given but cooperation is requested. The usual social constraints (i.e., desire for approval, fear of disapproval) provide an important source of motivation for subjects in whom $M_{af} > M_s$. If these subjects are to perform the task at all, some positive motivation must exist to oppose their tendency to avoid an achievement task. For subjects in whom $M_s > M_{af}$, the extrinsic motivation to perform an assigned task enhances their normally positive motivation to perform achievement related tasks. In both cases, task performance may be considered overdetermined, that is, the result of two or more different kinds of motivation to perform or not to perform the task (cf. Atkinson & Reitman, 1956).

In the light of this analysis of the components of total motivation, the basic condition for performance of the initial task rather than the alternative task is as follows:

INITIAL TASK				ALTERNATIVE TASK		
Achievement related motivation	+	Extrinsic motivation	>	Achievement related motivation	+	Extrinsic motivation

It is assumed that the subject will turn to the alternative achievement task when:

INITIAL TASK				ALTERNATIVE TASK		
Achievement related motivation	+	Extrinsic motivation	<	Achievement related motivation	+	Extrinsic motivation

How then does total motivation to perform the initial achievement task become weaker than total motivation to perform the alternative, so that the subject gives up persisting at the initial task and turns to the alternative? To permit derivation of hypotheses the following three assumptions are made:

1. Both extrinsic motivation to perform the initial task and extrinsic motivation to perform the alternative task are assumed to be constant across experimental conditions.

2. Extrinsic motivation to perform the initial task is assumed to be stronger than extrinsic motivation to perform the alternative task. This is a strong assumption that at least acknowledges the fact that the subject is asked by the experimenter to begin the initial task first.

3. The subjective probability of success for the alternative achievement task is assumed to be constant across experimental conditions. Hence, achievement related motivation to perform the alternative task is a positive constant for subjects in whom the motive to achieve success is stronger (i.e.,

when $M_s > M_{af}$), and has the same value (but negative) for subjects in whom $M_{af} > M_s$.

It follows that any decrease in total motivation to perform the initial task must be related to change in achievement related motivation to perform the initial task. Since, in the theory of achievement motivation, the motives to achieve success (M_s) and to avoid failure (M_{af}) are considered relatively *stable* dispositions of the personality, change in achievement related motivation to perform the initial task must be mediated by variation in the subject's expectation of success at the task as he works at it unsuccessfully. Repeated unsuccessful attempts at the initial task are assumed to produce successive decreases in subjective probability of success at the task (P_s). Decrease in P_s is thus the basic dynamic principle assumed to mediate decrease in total motivation to perform the initial task.

The following two specific assumptions are made about this decrease in expectation of success:

1. When the task is presented as a test of skill, reduction in P_s to a particular value is assumed to require more unsuccessful attempts at the task when P_s is initially high than when P_s is initially low.[4]

2. The rate at which decrease in P_s occurs is assumed not to be systematically related to the strength of either the motive to achieve success (M_s) or the motive to avoid failure (M_{af}).

Hypotheses

The following four hypotheses are tested in the present investigation:

Hypothesis 1 states that when the motive to achieve success is stronger than the motive to avoid failure ($M_s > M_{af}$), persistence at the initial achievement task should be greater when initial subjective probability of success is high (i.e., some $P_s > .50$) than when initial P_s is low (i.e., some $P_s < .50$).

Hypothesis 2 states that when $M_{af} > M_s$, persistence at the initial achievement task should be greater when initial P_s is low (i.e., some $P_s < .50$) than when initial P_s is high (i.e., some $P_s > .50$).

[4] This assumption is consistent with the distinction between tasks involving skill and tasks involving a degree of external control (e.g., chance) presented by James and Rotter (1958).

Hypothesis 3 states that when initial P_s is high (i.e., some $P_s > .50$), subjects in whom $M_s > M_{af}$ should persist longer at the initial achievement task than subjects in whom $M_{af} > M_s$.

Hypothesis 4 states that when initial P_s is low (i.e., some $P_s < .50$), subjects in whom $M_{af} > M_s$ should persist longer at the initial achievement task than subjects in whom $M_s > M_{af}$.

The detailed derivation is presented for Hypotheses 1 and 2 only. The derivation of Hypotheses 3 and 4 follows the same line of argument.

Derivation of Hypothesis 1

Consider a subject in whom the motive to achieve success is stronger than the motive to avoid failure (i.e., $M_s > M_{af}$), and to whom the initial achievement task is presented as easy (i.e., some $P_s > .50$). For this subject, achievement related motivation to perform the task is *positive* (approach) and becomes increasingly positive as his subjective probability of success (P_s) at the task falls to .50 with successive failures. Hence, total motivation to perform the initial task should increase as the task appears more difficult to the subject up to the point at which his $P_s = .50$. At this point positive achievement related motivation and, hence, total motivation to perform the task would be maximum. Thereafter, as P_s gets lower and the task appears more and more difficult to the subject, positive achievement related motivation decreases until at some low level of P_s total motivation to perform the initial task becomes weaker than total motivation to perform the alternative. The subject should then quit the initial task and turn to the alternative.

In contrast, if the initial task were presented to this subject as very difficult (i.e., some $P_s < .50$), there would be an immediate decrease in the positive achievement related motivation to perform the task as his P_s drops with repeated failure. Hence, total motivation to perform the initial task would decrease immediately. In this condition it should take less unsuccessful attempts at the task to reduce P_s to the low level at which total motivation to perform the initial task becomes less than total motivation to perform the alternative. Hence, we would expect a subject in whom

$M_s > M_{af}$ to persist longer at the initial achievement task when his P_s is initially high than when his P_s is initially low.

Derivation of Hypothesis 2

Consider a subject in whom the motive to avoid failure is stronger than the motive to achieve success (i.e., $M_{af} > M_s$), and to whom the initial achievement task is presented as easy (i.e., some $P_s > .50$). For this subject, achievement related motivation to perform the task is *negative* (avoidance) and becomes increasingly negative as his P_s at the task drops to .50 with repeated failure. Hence, total motivation to perform the initial task should decrease as the task appears more difficult to the subject up to the point at which his $P_s = .50$, where negative achievement related motivation would be maximum. At this point total motivation to perform the initial task should therefore be minimum. It follows that, if the subject is to quit the task at all, he should do so as his P_s at the task falls to .50 since it is during this stage of task performance that total motivation to perform the task is decreasing.

In contrast, if the initial task were presented to this subject as very difficult (i.e., some $P_s < .50$), there would be an immediate decrease in negative achievement related motivation as his P_s drops with repeated failure. Hence, total motivation to perform the initial task would increase immediately, and the subject should continue to perform the task indefinitely. We would therefore expect a subject in whom $M_{af} > M_s$ to persist longer at the initial achievement task when his P_s is initially low than when his P_s is initially high.

In the preceding analysis, differences in persistence have been related to differences in the way in which total motivation to perform the initial achievement task changes with decrease in P_s as a subject experiences repeated failure at the task. The different types of change in total motivation are summarized in Table 1 in relation to differences in motive and differences in initial P_s.

Adequate tests of the hypotheses of the study clearly require development of procedures which communicate to the subjects that they are expected to perform the initial achievement task first, but that they may feel free to move on to the alternative achievement task whenever they choose to do so. If extrinsic motivation to perform the initial task were very strong in relation to extrinsic motivation to perform the alternative task, none of the differences in persistence predicted here would be expected to occur.

TABLE 1

CHANGES IN TOTAL MOTIVATION TO PERFORM THE INITIAL ACHIEVEMENT TASK AS P_s DECREASES IN RELATION TO DIFFERENCES IN MOTIVE ($M_s > M_{af}$ and $M_{af} > M_s$) AND DIFFERENCES IN INITIAL SUBJECTIVE PROBABILITY OF SUCCESS ($P_s > .50$ and $P_s < .50$)

Initial P_s	Motive relationship	Change in total motivation as P_s decreases
$> .50$	$M_s > M_{af}$	Increase followed by decrease
$< .50$	$M_s > M_{af}$	Immediate decrease
$> .50$	$M_{af} > M_s$	Decrease followed by increase
$< .50$	$M_{af} > M_s$	Immediate increase

METHOD

Subjects

Thematic apperceptive stories under neutral conditions and Mandler-Sarason Test Anxiety Questionnaire provided measures of strength of n Achievement (M_s) and Test Anxiety (M_{af}) for 89 male college students from an introductory psychology course at the University of Michigan in 1959.[5] The TAT was administered according to the standard procedure (McClelland, Atkinson, Clark, & Lowell, 1953). Six pictures were presented in the following order (using numbers assigned by Atkinson, 1958): 2, 48, 1, 7, 100, 24. Interscorer reliability for scoring n Achievement was .93.[6]

In subjects classified High n Achievement-Low Test Anxiety (in terms of median splits) it was assumed $M_s > M_{af}$; in subjects classified Low n Achievement-High Test Anxiety it was assumed $M_{af} > M_s$.

Task

A task consisting of four items and labeled the Perceptual Reasoning test was presented to 34 of these preselected subjects in individual test sessions of one hour duration. Subjects were randomly assigned so that approximately half of each of the High n Achievement-Low Test Anxiety and Low n Achievement-High Test Anxiety groups came to the first item of the Perceptual Reasoning test with a high initial subjective probability of success (P_s); the other half came to the first item with a low initial P_s. The procedure adopted in assigning subjects guaranteed that the experimenter would be unaware when testing a subject for persistence of his n Achievement and Test Anxiety scores. The initial P_s for

[5] The author wishes to thank Charles P. Smith for administering the TAT and Test Anxiety Questionnaire.

[6] The author wishes to thank William Larkin for his assistance in scoring the TAT protocols.

each item of the test was experimentally induced by using fictitious group norms and the subject could leave each item as he wished and turn to the next one in the series.

Each item of the Perceptual Reasoning test involved a line diagram approximately 1.5 inches square printed on a white card 6″ × 4″. Copies of a particular item were stacked in a pile 2″ high and the four piles of cards were placed in line on the desk immediately in front of the subject. The piles were designated 1–4 from left to right and the subject could not see the content of any item until he began to work at it, i.e., the precise content of the alternatives to which he could turn if he wished to quit an item was unknown to him.

For each item the subject's task was to trace over all the lines of the diagram with a red pencil according to two rules: he was not permitted to lift his pencil from the figure; he was not permitted to trace over any line twice. It is a simple matter to construct figures for which the problem is *insoluble* but which are sufficiently complex so that the subject is unable to see this. The subject was allowed to take as many trials at an item as he wished. If he wanted another trial at the item, he simply took another copy of the same item from the top of its pile and worked at it. The following instructions to the subject summarize the restrictions placed on task performance:

> Naturally there are certain restrictions involved in this test. You can only work at an item for 40 seconds at a time, and I shall be timing you on that. But you can have as many of these 40 second trials as you want. If you fail on a trial, that is if you don't succeed in tracing over all the lines in the figure, you will then have the choice of continuing or going on to the next item. If you want to go on to the next item you should let me know at once. Once you've stopped working at an item you can't go back to it again. If you want to continue with the item you should turn the *failed* copy face downwards. You can then take another copy of the same item from the pile and you will again have 40 seconds to work at that.

Obviously, for an insoluble item, each trial the subject took resulted in failure. Using this procedure, one can obtain two measures of persistence for an insoluble item: total time spent in working at the item; and number of trials taken at the item before turning to the next one in the series. The former measure is akin to the typical temporal measure used in studies of persistence; the latter is analogous to resistance to extinction.

It will be noted that the alternative activity to which the subject could turn was another *achievement* task similar in content to the one at which he had been working.

Of the four items involved in the Perceptual Reasoning test the first and third were *insoluble*, the second and fourth items were soluble.

Experimental Induction of Subjective Probability of Success

The following instructions were given:

> Now the four items vary in difficulty. Some are harder than others, and you're not expected to be able to solve all of them, but do the best you can. Before I present each item I'm going to let you know the percentage of college students who are able to pass that item. . . . Well, let's look at the norms for the first item. These go by age. How old are you? The tables show that at your age level——per cent, that is approximately——per cent of college students are able to get the solution.

Table 2 presents the fictitious norms for each of the four items of the Perceptual Reasoning test that were reported to Groups A, B, C, and D. These norms were assumed to be sufficiently different to determine differences in initial P_s for each of the four items. It should be noted that initial P_s on items following the first may be influenced not only by the norms reported to the subject but also by the subject's actual experience with the preceding items.

The following final instructions were given before Item 1 was presented to the subject:

> Try to get the solution if you can. It's quite OK for you to take as many 40 second trials as you want at the item. But remember that the four items do vary in difficulty, some are harder than others and you may not be able to solve all of them. So if you should feel that you're not getting anywhere with the item you should let me know at once so that we can move on to the next one. Here is Item 1. You should find it fairly easy/difficult.

For each insoluble item the subject was timed from when he started the item until when he nominated to try the next item in the set. When he quit an insoluble item, the copies of the item he had attempted were stacked to one side out of sight and their number was later recorded.

Effectiveness of Reported Norms

Norm procedure: After reporting the fictitious group norm but before presenting the item, the experimenter gave the following instruction: ". . . I'm also going to ask you to let me know what you think your own chances are of solving the problem. You can do this very simply by marking a number along this scale. I'll get you to do this before I present each item and I'll also get you to check here how confident you are about that estimate." To obtain the probability estimates a 20-point rating scale was used, numbered from 0 to 100

TABLE 2

FICTITIOUS PERCENTAGES[a] OF SUCCESS REPORTED TO SUBJECTS IN THE EXPERIMENTAL INDUCTION OF INITIAL DIFFICULTY FOR EACH OF FOUR TEST ITEMS

Group	Motivation		Reported percentages of success for each of four items			
	n Achievement	Test Anxiety	1	2	3	4
A	High	Low	70	50	5	50
B	High	Low	5	50	70	50
C	Low	High	70	50	5	50
D	Low	High	5	50	70	50

[a] Actual percentages reported deviated slightly from tabulated values in that they were given to one decimal place to add to authenticity (e.g., 69.4%).

in steps of five. The subject made his rating of confidence in his probability estimate on a five-category Likert-type scale ranging from "Not certain at all" to "No doubt about it at all."

Time Limits for Insoluble Items 1 and 3

Data concerning persistence at Item 1 were from the very outset considered more crucial for the test of the hypotheses since results for other items would be complicated by differential sequence effects. For this reason each subject was allowed to persist at Item 1 for 20 minutes (if he so desired) so as to increase the possibility of obtaining a wide variance in persistence scores. If the subject was still persisting at Item 1 after 20 minutes it was suggested that he move on to Item 2. This interruption procedure for Item 1 necessarily introduced a time differential for Item 3, some subjects having more time to work at this item than others, depending on how long they had worked at Item 1. Subjects still working at Item 3 were interrupted at approximately 15 minutes before the hour to allow sufficient time for administration of a postperformance questionnaire.

Postperformance Questionnaire

Finally, the experimenter administered a questionnaire to each subject specifically designed to provide information about the assumptions involved in the hypotheses. The majority of questions were presented in Likert-type form and were related to Item 1. When the subject had completed the questionnaire he was thanked for his participation in the experiment, commended on his performance at a fairly difficult task, and requested to keep details of the experiment in confidence. A report of the experiment was sent to each subject at a later date.

RESULTS

The basic data relevant to the hypotheses of the present investigation are the persistence scores of the four groups in terms of time and trials for the first insoluble item (Item 1). The results of analysis of persistence scores for the second insoluble item (Item 3) are presented in less detail since the necessity for unanticipated interruption of a sizable number of extremely persistent subjects on the first and third items, and the possibility of uncontrolled sequence effects, make interpretation of results for the third item equivocal.

Analysis of Persistence at Item 1

Since time and trials scores are perfectly correlated for Item 1, only the analysis of persistence trials is presented. By persistence trials is meant the number of trials the subject took at Item 1 before he quit to move on to the second item or before he was interrupted. The median of these persistence trials at Item 1 for all subjects is 20 with a range from 2 to 41+.[7] Persistence trials for subjects of each of Groups A, B, C, and D are classified as high and low in terms of whether they are above or below this median, and the resulting frequencies are presented in Table 3. Below Table 3 are presented the results of a partitioned χ^2 analysis (Sutcliffe, 1957) applied to these data.

Table 3 shows that only the χ^2 representing the interactive effect of motivation and expectation on persistence is significant ($p < .01$). While the proportion of subjects classified high in persistence *increases* as the reported norm varies from 5% to 70% (i.e., from difficult to easy) for subjects high in n Achievement and low in Test Anxiety, this proportion *decreases* for subjects low in n Achievement and high in Test Anxiety with the same variation in the reported norm. This implies a basic difference in the way persistence is related to the initial difficulty of the item among subjects in whom $M_s > M_{af}$ versus subjects in whom $M_{af} > M_s$.

Relationships in Table 3 are clearly consistent with Hypotheses 1, 2, 3, and 4. Subjects in whom it is assumed that $M_s > M_{af}$ (the High-Low subjects) show greater persistence at the task when the reported norm defines it as easy (70% norm) than when the norm defines it as difficult (5% norm). This result supports Hypothesis 1 in implying that subjects in whom $M_s > M_{af}$ tend to persist longer at the item when initial expectation of success is high rather than low. In contrast, subjects in whom it is assumed that $M_{af} > M_s$ (the Low-High subjects) show greater persistence at the task when the reported norm defines it as difficult (5% norm) than when the norm defines it as easy (70% norm). This result supports Hypothesis 2 in implying that subjects in whom $M_{af} > M_s$ tend to persist longer at the item when initial expectation of success is low rather than high. Furthermore, among subjects for whom the reported norm defines the item as easy, the proportion of subjects showing high persistence is greater for those classified as high

[7] The + indicates that the subject had to be interrupted. Due to the interruption procedure some scores are open-ended. Hence, results are classified on the basis of median splits and relationships in the resulting frequency tables analyzed using χ^2 as the basic statistic.

TABLE 3

NUMBER OF SUBJECTS WHO WERE HIGH AND LOW IN PERSISTENCE ON ITEM 1 IN RELATION TO STATED DIFFICULTY OF THE TASK AND NATURE OF MOTIVATION

Group	Motivation		Stated difficulty of task	Persistence trials	
	n Achievement	Test Anxiety		High (above median)	Low (below median)
A	High	Low	70% (easy)	6	2
B	High	Low	5% (difficult)	2	7
C	Low	High	70% (easy)	3	6
D	Low	High	5% (difficult)	6	2
A and B	High	Low	70%, 5%	8	9
C and D	Low	High	70%, 5%	9	8
A and C	Both conditions		70% (easy)	9	8
B and D	Both conditions		5% (difficult)	8	9

Partition of χ^2

Source	Value	*df*	*p*
Motivation × Persistence	.12	1	> .05
Expectation × Persistence	.12	1	> .05
Motivation × Expectation × Persistence	7.65	1	< .01
Total	7.89	3	< .05

in n Achievement and low in Test Anxiety. This result is consistent with Hypothesis 3 in implying that, when the initial expectation of success is high, subjects in whom $M_s > M_{af}$ tend to persist longer at the item than subjects in whom $M_{af} > M_s$. Finally, among subjects for whom the reported norm defines the item as difficult, the proportion of subjects showing high persistence is greater for those classified as low in n Achievement and high in Test Anxiety. The result is consistent with Hypothesis 4 in implying that, when the initial expectation of success is low, subjects in whom $M_{af} > M_s$ tend to persist longer at the item than subjects in whom $M_s > M_{af}$.

Analysis of Persistence at Item 3

Persistence data for Item 3 compared across Groups A, B, C, and D show no significant trends. However, when comparisons of persistence scores for Items 1 and 3 are made within the same group, subjects of Group B (High-Low subjects) persist longer at Item 3 which for them is defined by the reported norm as easy than at Item 1 which is defined by the norm as difficult. This change is consistent with Hypothesis 1. Subjects of Group C (Low-High subjects) persist longer at Item 3 which for them is defined by the reported norm as difficult than at Item 1 which is defined by the norm as easy. This change is consistent with Hypothesis 2. However, when ambiguous comparisons of persistence scores are excluded (due to the interruption procedure), there is a tendency for all subjects to persist longer at Item 3 than at Item 1. This suggests that an influence common to all subjects might have operated to increase persistence at Item 3.

When such an influence is assumed and ambiguous differences in persistence scores (due to the interruption procedure) are excluded, the data support the qualified prediction that Groups A and D should at least tend to increase relatively less in persistence from Item 1 to Item 3 than do Groups B and C. However, this result is equivocal since, as a result of limitations of time, subjects of Group A and D who persisted relatively longer at Item 1 may not have had the same opportunity to increase in persistence at Item 3 as did subjects of Groups B and C.

Analysis of Supplementary Data

Analysis of data concerned with effectiveness of experimental procedures supports the assumption that the different reported norms determined widely different initial subjective probabilities of success. The probability estimates given by subjects in all four groups tend to follow the reported norms very closely.

Data obtained from the postperformance questionnaire are generally consistent with assumptions involved in hypotheses. For example, subjects in all four groups report decreases in estimates of probability of success as they work at Item 1 with repeated failure. This decrease is consistent with the basic dynamic assumption that expectation of success should diminish with repeated failure at the task. In addition, there is no evidence that the rate of decrease in probability estimates with failure at Item 1 differs between the two motivation conditions (High-Low subjects versus Low-High subjects). This implies that rate of decrease in P_s is independent of the relative strength of M_s and M_{af} within the individual.

DISCUSSION

Where the present investigation permits unequivocal test of the four hypotheses (i.e., for Item 1), results are consistent with predictions. Furthermore, postperformance questionnaire data generally support the basic assumptions involved in these predictions.

The investigation stands in marked contrast to the typical trait studies of persistence and extinction studies with human subjects since hypotheses are based on a consideration of personality and situational factors in interaction. Trait studies of persistence typically concentrate on the relatively stable aspects of persons by looking for consistencies in behavior and they tend to neglect the role of the situation. In contrast, extinction studies usually attempt to relate persistence of behavior to characteristics of the acquisition stage and they tend to ignore the influence on resistance to extinction of relatively stable personality characteristics. The contribution of the present study is to investigate persistence as a phenomenon determined by the interaction of both personality and situational influences.

The present investigation also differs from the usual extinction investigation by using a task that could be seen by the subject as involving personal skill (cf. James & Rotter, 1958) and that provided a relatively inexhaustible range of alternative responses following failure. In contrast, extinction studies with humans have often used tasks in which the subject may see success and failure as largely beyond his control (e.g., the Humphrey's apparatus) and where the response is restricted to a particular action (e.g., pulling down a lever). Furthermore, in the present investigation initial expectation of success was varied by using fictitious norms. In contrast, in extinction studies the expectation may be considered as based mainly on frequency and pattern of reinforcement and nonreinforcement in the acquisition series.

The theoretical analysis may be extended to a consideration of persistence under conditions that differ from those of the present investigation in important respects. The general paradigm of the persistence situation is that in which the subject has some expectation of attaining a goal, is unrestricted in the number of attempts he can have at attaining it, and has an alternative activity to turn to as he wishes. Within this general framework the following differences may be noted which in turn imply specific types of persistence situation worthy of investigation:

1. Performance of the alternative to which the subject may turn may involve motives that are the same or different from those involved in the performance of the activity at which he is presently engaged. The persistence situation studied in the present investigation is one in which initial task and alternative activity belong to the same class of activity. Total motivation to perform the initial task and total motivation to perform the alternative task are both attributable to the *same* motivational components. In contrast, we can consider a type of persistence situation in which the initial task and the alternative task do not belong to the same class of activity, i.e., where the alternative is *motivationally different* in the sense that total motivation to perform the alternative is influenced by motives which are not involved in performance of the initial activity. Investigations of persistence by French and Thomas (1958) and by Atkinson and Litwin (1960) appear to fall in this latter class since in both of these studies the alternative activity to which the subject can turn does not necessarily involve achievement related motivation.

2. The situation does or does not involve types of motivational tendency in which incentive values and expectations are assumed to be dependent. In the achievement context

dependencies are assumed between the positive incentive value of success and the subjective probability of success, and between the negative incentive value of failure and the subjective probability of failure (cf. Feather, 1959a, 1959b). But, in the more general case (e.g., efforts to get food when hungry), there would be no dependency between incentive value and expectation, and, hence, no increase to a maximum motivation followed by a decrease as expectation of goal attainment (P_g) changes from $P_g > .50$ to $P_g < .50$ (as occurs for achievement related motivation). Instead we would expect motivation to perform the activity to decrease continuously with decrease in the strength of expectation of goal attainment, providing strength of motive and incentive remained constant.

3. Persistence situations may differ according to whether or not the incentive is objectively present. One can consider a persistence situation in which the incentive is actually in view, e.g., young children attempting to overcome an obstacle in order to get a candy. In the present investigation and in the human extinction type of study, the incentive is not objectively present. In the Perceptual Reasoning test, the subject has to solve the puzzle; he cannot see the solution. In the extinction situation the reinforcement does not appear in the extinction series. Hence, it is possible for the subject to develop the expectation that there is in fact no incentive, e.g., that the puzzle is insoluble or that the experimenter has "fixed" the apparatus so that no more reinforcements will occur. In other words, the subject may conceive of these situations as beyond his control, as not involving skill. Consequently, special care is required to ensure that the subject retains the belief that only his efforts will influence the outcome.

These three differentiating characteristics, in combination, yield the eight specific types of persistence situation summarized in Table 4. The persistence situation investigated in the present study corresponds to 5, i.e., performance of the alternative involves the same kind of motivation as the activity at which the subject is presently engaged, dependencies are assumed between incentive values and expectations, and the incentive is not objectively present. In contrast, the

TABLE 4

TYPES OF PERSISTENCE SITUATION ACCORDING TO WHETHER OR NOT THE ALTERNATIVE IS MOTIVATIONALLY THE SAME OR DIFFERENT FROM THE PRESENT ACTIVITY, DEPENDENCIES BETWEEN INCENTIVE VALUES AND EXPECTATIONS EXIST OR DO NOT EXIST, AND THE INCENTIVE IS OBJECTIVELY PRESENT OR ABSENT

Persistence situation	Alternative (motivationally same or different)	Dependency between expectation and incentive value (present or absent)	Incentive (objectively present or absent)
1	same	present	present
2	different	present	present
3	same	absent	present
4	different	absent	present
5	same	present	absent
6	different	present	absent
7	same	absent	absent
8	different	absent	absent

persistence situation studied by French and Thomas (1958) and by Atkinson and Litwin (1960) corresponds to 6 since performance of the alternative in these studies probably involves different motivation from that involved in the activity at which the subject is engaged. It should be possible to construct experimental situations corresponding to the other six cases and to analyze persistence behavior in each of these in terms of the interaction of motives, expectations, and incentives. Such an approach would require (*a*) specification of the nature of the motives, expectations, and incentives involved in the situation; (*b*) measures of the strength of these motives, expectations, and incentives; (*c*) assumptions concerning how these basic variables combine to determine the various component motivations; (*d*) specification of the manner in which expectation changes with experience at the activity; and (*e*) assumptions concerning the resolution of the set of component motivations into a resultant motivation.

The present study has shown that this strategy does work for the type of persistence situation investigated here. Detailed specification of the components of total motivation to perform a task in relation to total motivation to perform the available alternative, in conjunction with a dynamic principle of expectation change, has permitted differ-

ential predictions about persistence. This type of analysis, which takes account of both personality dispositions and situational influences in interaction according to an "expectancy value" model, may also be required to explain persistence in the other types of situation.

SUMMARY

This investigation examines the relationship of persistence at a task both to its apparent difficulty and to the relative strength within an individual of the motives to achieve success and to avoid failure. A subject performs a task presented to him as part of an important test, undergoes repeated failure at it (because it is in fact insoluble), but may turn to an alternative achievement task whenever he desires. Strength of motive to achieve success is assessed by the TAT n Achievement procedure; strength of motive to avoid failure is assessed by Mandler-Sarason Test Anxiety Questionnaire. Apparent difficulty of the task (i.e., initial expectation of success) is varied by use of fictitious group norms. Persistence is measured by total trials or total time the subject works at the task before turning to the alternative.

Hypotheses are derived from Atkinson's theory of achievement motivation with additional assumptions. The following hypotheses are tested:

Hypothesis 1 states that when motive to achieve success (M_s) is stronger in the subject than motive to avoid failure (M_{af}) he should be more persistent at a task when his initial subjective probability of success (P_s) is high (easy task, $P_s > .50$) than when his initial P_s is low (difficult task, $P_s < .50$).

Hypothesis 2 states that when $M_{af} > M_s$, the subject should be more persistent when his initial P_s is low than when his initial P_s is high.

Hypothesis 3 states that when initial P_s is high, a subject in whom $M_s > M_{af}$ should be more persistent than one in whom $M_{af} > M_s$.

Hypothesis 4 states that when initial P_s is low, a subject in whom $M_{af} > M_s$ should be more persistent than one in whom $M_s > M_{af}$.

Results are generally consistent with hypotheses and with assumptions involved in the predictions.

As a guide to future research a theoretical scheme is presented in which different classes of persistence situations are identified. The present investigation indicates the importance of specifying the components of total motivation to perform a task in relation to the components of total motivation to perform the alternative when attempting to predict degree of persistence. The theoretical analysis suggests the possibility of conceptualizing each component motivation in "expectancy value" terms, and the importance of change in expectation as a dynamic principle mediating change in motivation. The factual evidence clearly provides a demonstration that persistence can be conceptualized as an interaction of personality dispositions and situational influences.

REFERENCES

ATKINSON, J. W. Motivational determinants of risk-taking behavior. *Psychol. Rev.*, 1957, **64,** 359–372.

ATKINSON, J. W. (Ed.) *Motives in fantasy, action, and society.* Princeton: Van Nostrand, 1958.

ATKINSON, J. W. Personality dynamics. *Annu. Rev. Psychol.*, 1960, **11,** 255–290.

ATKINSON, J. W., & LITWIN, G. H. Achievement motive and test anxiety as motives to approach success and to avoid failure. *J. abnorm. soc. Psychol.*, 1960, **60,** 52–63.

ATKINSON, J. W., & RAPHELSON, A. C. Individual differences in motivation and behavior in particular situations. *J. Pers.*, 1956, **24,** 349–363.

ATKINSON, J. W., & REITMAN, W. R. Performance as a function of motive strength and expectancy of goal attainment. *J. abnorm. soc. Psychol.*, 1956, **53,** 361–366.

FEATHER, N. T. Subjective probability and decision under uncertainty. *Psychol. Rev.*, 1959, **66,** 150–164. (a)

FEATHER, N. T. Success probability and choice behavior. *J. exp. Psychol.*, 1959, **58,** 257–266. (b)

FEATHER, N. T. Persistence in relation to achievement motivation, anxiety about failure, and task difficulty. Unpublished doctoral dissertation, University of Michigan, 1960.

FRENCH, ELIZABETH G. Some characteristics of achievement motivation. *J. exp. Psychol.*, 1955, **50,** 232–236.

FRENCH, ELIZABETH G., & THOMAS, F. H. The relation of achievement motivation to problem solving effectiveness. *J. abnorm. soc. Psychol.*, 1958, **56,** 46–48.

JAMES, W. H., & ROTTER, J. B. Partial and 100% reinforcement under chance and skill conditions. *J. exp. Psychol.*, 1958, **55,** 397–403.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

SUTCLIFFE, J. P. A general method of analysis of frequency data for multiple classification design. *Psychol. Bull.*, 1957, **54,** 134–137.

(Received November 22, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 63, No. 3, 562-569

IMPULSIVE CLOSURE AS REACTION TO FAILURE-INDUCED THREAT[1]

JAMES E. DITTES
Yale University

CONFRONTING diffuse, disorganized, and ambiguous pieces of information in the environment, persons vary in the way in which they impose conclusions, patterns, or meaning. They vary not only in the nature of the closure they impose, but also in their readiness to impose any closure at all. Given a particular series of stimuli, some persons form a closure readily as best they can with the available information, and others tend to defer judgment, leaving the stimuli ambiguous until more information is available. In most instances, a moderate delay in closure is more adaptive and produces a higher quality closure than does an impulsive judgment.

Such variability in the tendency to form a closure has not been widely recognized and studied, except in special cases such as those labeled as rigidity, conformity, persuasibility, suggestibility, ethnocentrism, prejudice, and resistance to free association. These special cases attract interest either because the behavior is maladaptive or because particular identifiable sources are used by the person to help him form the closure. Examples of such identifiable sources are the judgments of others in the case of conformity, or one's own prior judgments, in the case of rigidity.

So far as individual differences in the tendency to closure have been studied, they have generally been attributed to cognitive impairment due to distraction or emotional interference (e.g., Asch, 1952; Moffitt & Stagner, 1956; Shipley, 1956), or else to different "styles" (Frenkel-Brunswik, 1949; Gardner, Holzman, Klein, Linton, & Spence, 1959; Holzman, 1954; Klein, 1958; Messick & Hills, 1960). The latter line of investigation has looked for correlated patterns in various aspects of cognitive and related behavior.

A motivational basis for differences in the strength of tendency to impose closure has been proposed (Dittes, 1959) with the theory that closure, regardless of its contents, acquires a specific reward value (in most sectors of our culture at least), as a means of enhancing self-esteem and self-confidence. The reward value of closure is the result of previous experience in a wide variety of circumstances in which important need satisfactions are dependent upon or closely associated with cognitive closure. Examples would include parental approval of a child's learning to name objects or the relief following recognition of a pattern of job promotion which makes it clear that being passed over is not due to personal inadequacies. On the basis of such acquired reward value, the prediction follows that persons experiencing a greater need to maintain or enhance their sense of self-esteem are more likely to impose closure.

A previous experiment has given support to this general theory (Dittes, 1959). Self-esteem was manipulated via apparent acceptance by peers, social approval presumably being one important source of self-esteem. It was found that persons made to experience low approval tended subsequently to demonstrate more impulsive closure than those who experienced higher approval.

However, the results of this single experiment could be attributed not to the mediation of threatened self-esteem but more directly to the variable of social acceptance, persons seeking to impose closure as a kind of "show-off" because they believed this would win them regained acceptance. This interpretation of the previous results remained possible, although it was made unlikely by several controls that were used.

The self-esteem theory predicts more generally that similar results could be found when self-esteem is manipulated by some quite different variable. The present experiment undertook to affect self-esteem by manipulating the subjects' sense of success or failure on an ego involved task. Care was taken to rule out any possible effects of actual or anticipated social approval. Performance was anonymous so far as the subjects knew and the subjects could not anticipate any social approval

[1] This experiment was supported by Grant M-2590 from the National Institute of Mental Health, United States Public Health Service.

or disapproval because of their performance. A sample of graduate students was used to provide subjects who were likely to be particularly vulnerable to success and failure manipulations.

An experimental group was made to experience failure on an ego involved task. Two control groups were used: one that was made to experience success, and one that experienced failure to the same degree as the experimental group but was led to regard the task as unrelated to crucial self-esteem factors. The two control groups were required to isolate the effects of esteem-threatening failure.

The same measures of impulsive closure were used as in the preceding experiment and the same measures of personality characteristics were also used with two additions. The specific hypotheses were as follows:

Effect of Failure-Induced Threat

Failure produces more impulsive closure than either success or failure on a noninvolving task.

Effects of Personality Characteristics

Self-esteem. In a given situation without immediate strong situational support of self-esteem, persons with lower general self-esteem show more impulsive closure than those with higher general self-esteem. Aside from situational influences on self-esteem that may affect closure, persons with lower general self-esteem are presumably constantly seeking whatever defensive supports may be available and are more likely to resort to closure, when circumstances provide opportunity than persons with higher general self-esteem. Differences attributed to general personality characteristics are expected only in the experimental condition and not in the control conditions, because in the latter, self-esteem for all persons is at least temporarily elevated. Differences attributable to personality characteristics that persons bring with them into the experiment are likely to show up only when the situation is depriving of immediate support.

Valuation of achievement. In a situation in which achievement motivation is frustrated, persons who attach strong value to achievement show more impulsive closure than persons weak in achievement valuation. Persons with stronger achievement motives presumably are more threatened by such frustration and therefore more motivated to seek compensating satisfaction, such as closure.

Interaction of Threat and Personality

Self-esteem. The predicted effects of failure are likely to occur only among persons with low general self-esteem. These are the persons who are vulnerable to such manipulation. Persons with high self-esteem presumably have other resources and are less threatened, if at all, by a single situation of failure.

Valuation of achievement. The predicted effects of failure are likely to occur only among persons with high value on achievement. The same reasoning applies. These are the persons who are vulnerable to such manipulation. Persons who place less value on achievement presumably depend upon other resources for their self-esteem and would be vulnerable to the manipulation of them but not to the manipulation of achievement.

METHOD

Subjects

The experiment was conducted, without prior announcement, during a regular class meeting of a required course at the Yale Divinity School. Fifty-seven students completed the experimental materials.

To minimize any effects of interpersonal interaction and anticipated opinion of others, all work was anonymous, so far as the subjects knew, and all materials were self-administered, after a brief general introduction by the instructor. The experimental materials, which were distributed at one time in a folder, were represented as a series of unrelated tasks at which subjects were asked to work privately and anonymously. Assignment to experimental conditions was determined by the way in which the shuffled set of folders chanced to be distributed.

Experimental Manipulations

Subjects took Items 1–10, 31–40 of the Space Relations Test published by the Psychological Corporation. This test is difficult enough to prevent subjects from having much subjective certainty about the quality of their performance and it is therefore open to manipulation of reported performance results, and thus of success or failure. The test is also abstract enough to allow it to be represented as measuring anything; this provided opportunity for the ego involving manipulation. A special answer sheet was devised, so that answers were recorded, by carbon, onto a keyed sheet so as to permit self-scoring after the seal was broken. Anonymity was emphasized by repeated written instructions during the procedures that subjects should not write names on their materials and that the experimenter's interest was only in group averages not individual results.

Ego involving manipulation. Subjects assigned to the *success* and the *ego involved failure* conditions found written instructions describing the Space Relations Test as one measuring, "skills of abstraction . . . of central importance in almost all occupations, especially in the professions . . . highly related to personal effectiveness and professional success." Subjects who chanced to be assigned to the *noninvolved failure* condition found written instructions describing the test as "valid in predicting professional success with only two occupations—stagehands and certain kinds of architects" but indicating that for exploratory purposes it was desired to discover the average level of performance among the present sample.

The effectiveness of these manipulations is clearly indicated by the first two lines of Table 1. On a questionnaire administered at the end of the experiment subjects in the noninvolving condition reported trying significantly less hard to do well on the Space Relations Test than subjects in the ego involving conditions reported. Interestingly enough, their actual performance on the test was significantly superior, indicating presumably the fruits of their more relaxed approach. The second line of Table 1 reports the actual average performance on the Space Relations Test—as subsequently scored by the experimenter—to be distinguished from the results that subjects believed that they had attained.

Success-Failure manipulation. Fictitious keys and norms were used to affect the subjects' feelings of having performed well or poorly on the Space Relations Test. For subjects assigned to the success condition, a special answer key was devised, to maximize the number of apparently correct responses. A fictitious set of norms was provided so that all subjects would find their scores ranging among the high percentiles of graduate students. For subjects in the failure conditions, a false key was provided indicating that many of the most common responses were incorrect. A set of norms was also devised in such a way as to make all subjects range in the lower percentiles as compared with candidates for bachelor degrees.

To guarantee that subjects would inspect the norms carefully, a form was provided on which they were required to record their own score and percentile.

The middle section of Table 1 indicates that these manipulations were highly effective. On three items on the final questionnaire, success subjects reported that their results were above average, higher than they had expected, and pleasing, while subjects in both failure conditions reported their results were below average, lower than they expected, and disturbing.

Evidence of net effect on self-esteem. The pair of manipulations just described was intended to produce one condition (ego involving failure) producing a threat to self-esteem and two control conditions. Two items on the final questionnaire asked for the subjects' reports on this net effect, one a direct query and one indirect. The results, reported at the bottom of Table 1, seem to indicate that a satisfactory difference was achieved. Both nonthreat conditions reported that their results had been less on their minds through the subsequent part of the experiment than did the ego involved failure group. On the direct question, success subjects reported self-esteem mildly enhanced by performance on the Space Relations Test and ego involved failure subjects reported self-esteem mildly weakened. The noninvolved failure group reported an average score near the ego involved failure group but still substantially different ($p = .06$).

Measures of Impulsive Closure

Parable task. Subjects were presented with a one-paragraph story, written in Biblical idiom and represented as a portion of a recently discovered scroll, and asked to write their opinions concerning the meaning of the parable. The account was essentially incoherent

TABLE 1

EVIDENCE OF EFFECTIVENESS OF EXPERIMENTAL MANIPULATIONS

Manipulation and questionnaire item	Average score in each condition[a]				p
	Control groups		Threat group		
	Success (Ego involved)	Failure (Non-involved)	Failure (Ego involved)		
Ego involvement					
How hard did you try on test?	5.0	4.6	4.8	Noninvolved vs. Ego involved	$<.05$
Score on Space Relations Test	28.9	33.2	27.2		$<.05$
Success-Failure					
Test results above or below average?	2.4	5.8	5.7	Success vs. Failure	$<.001$
Test scores higher or lower than expected?	2.5	5.4	5.4		$<.001$
Pleased or disturbed by test results?	2.4	4.8	4.9		$<.001$
Net threat					
How much did test results stay in mind?	2.8	2.7	3.5	Control vs. Threat	.05
Test results affect self-esteem?	2.5	4.3	4.7		$<.001$

[a] Questionnaire responses based on six-point scales. Maximum possible score on this form of Space Relations Test was 54.

and unstructured. However, it contained a large number of familiar religious symbols that could, with some selection and distortion, be forced to yield a coherent meaning, though no two subjects ever found the same meaning. Responses were later scored on a three-point scale, according to whether each was a positive statement of meaning or indicated uncertainty as to meaning. The higher score represented more impulsive closure.

As for both tasks, the scoring categories were defined in advance of the experimental sessions, and the scoring was done by two judges working independently without knowledge of the subjects' performance on other tasks or of the experimental conditions to which the subjects had been assigned. The judges showed absolute agreement in 91% of these judgments.

This task was presented as entirely unrelated to the Space Relations Test. The rationale was offered that performance on this task would help discussion leaders in another course better prepare their sessions.

Roommate task. Subjects were given two lists of traits that were represented as words used by close friends to describe two students; they were instructed to write their impression of each man. Each list contained a dominant set of traits and a subordinate set that was incompatible with the dominant set. In one instance, for example, the list was: individualist, suspicious, belligerent, considerate, self-concerned thoughtful, stubborn, tolerant. The two lists were balanced so that dominant traits in one were favorable, in the other unfavorable.

Responses were judged on a three-point scale according to the extent to which the subordinate traits were integrated into the impression. A statement based exclusively on the dominant traits was regarded as the most impulsive. There was absolute agreement between the judges on 71% of the judgments. Differences between judges were averaged.

This task was represented as unrelated to other parts of the experiment, and as exploring one part of a procedure by which roommates might be selected.

Combined index.[2] Since both tasks were generated from the same rationale to measure a single behavioral tendency, the scores for each task were combined to form a more reliable over-all index of closure strength for each subject by summing the scores after weighting them by factors to equalize their variance.

Personality Measures

Several weeks before the experiment, subjects completed a self-rating questionnaire that included items comprising two different scales.

Self-esteem. This scale was comprised of 12 items such as:

I feel nervous if I have to meet lots of people.

I am usually not sure that others understand what I say.

Valuation on achievement. Six items assessed the general importance that subjects attached to personal achievement. Typical items were:

Successful accomplishment of a task is important whether others know of it or not.

Once I begin a task, I can't give up until I have mastered it.

Space relations aptitude. The subjects' actual performance on the Space Relations Test provided an estimate of general intellectual ability. This test is described by its publishers as measuring the "ability to manipulate things mentally," and correlates in the .50s with indices of general intelligence such as the ACE Psychological Examination.

Because of the false scoring key and norms used by the subjects, the actual test results are, of course, much higher and imperfectly correlated ($r = .82$) with the scores the subjects in the failure groups believed they had obtained. There is nevertheless the possibility that this measure is contaminated with the degree of failure experienced by subjects, those scoring lower perhaps experiencing more failure. However, with the false norms used by the subjects, all fell between 10th and 30th percentiles of college undergraduates so that there was relatively little room for individual differences in experienced failure, compared with the massive over-all effect. The subjectively experienced failure of a person falling at the 30th percentile should be at least as great as that of a subject scoring at the 10th percentile, because the former is presumably accustomed to higher scores.

Dichotomization of distributions. In order to assess the interaction between the experimental conditions and the personality characteristics, the scores on the distribution of each of these measures were dichotomized. As in the previous experiment, scores on the self-esteem distribution were divided into thirds and the higher third compared with the lower two-thirds. The other indices were divided at the median and subjects in the upper and lower halves of the distribution compared.

Table 2 presents the split-half reliabilities of these scales and the correlation among them. None of the three measures is significantly related with either of the others.

RESULTS

Effects of Failure-Induced Threat

Table 3 presents the impulsive closure scores on each of the tasks and for the combined index. The threatened subjects (ego involved failure) were more impulsive on each task than the subjects experiencing success. The noninvolved failure control group tended to be impulsive on the parable task but not on the roommate task, an inconsistent performance to be discussed later.

[2] The third task used in the earlier experiment (Dittes, 1959) proved to be unsuitable for large group administration, at least in the acoustics of the available room. This is the "jammed message" task which requires subjects to listen to a recorded tape. Performance with a large group was so highly related to position in the room as to swamp any possible differences due to experimental conditions.

Acoustics were probably most homogeneous for subjects seated in the back two-thirds of the room. Considering just this group, subjects in the ego involved failure condition performed more impulsively than subjects in the success condition ($p < .05$)

TABLE 2

INTERCORRELATIONS AND RELIABILITIES OF PERSONALITY MEASURES

	Intercorrelations		Reliability
	Valuation on achievement	Space Relations Test	
Self-esteem	.13	.01	.76
Valuation on achievement		−.20	.46

TABLE 3

AVERAGE IMPULSIVE CLOSURE SCORES

	Parable task	Roommate task	Combined index
Threat condition:			
Failure, ego involved	1.9	2.0	2.0
Control conditions:			
Success, ego involved	1.6	1.9	1.8
Failure, noninvolved	1.9	1.8	1.8
p Threat vs. Controls[a]	$<.05$	$<.05$	$<.05$

[a] Mann-Whitney U test (two-tailed) used throughout this report.

Effects of Personality Characteristics

Self-esteem. As predicted and as shown in the third column of Table 4, for the experimentally threatened condition low self-esteem subjects tended to show more impulsive closure than subjects with higher self-esteem. In the control condition no effects of differing self-esteem were predicted, because subjects were experiencing strong enough temporary support to make all subjects' self-esteem homogeneously high.

Valuation on achievement. Subjects with stronger valuation on achievement reacted to the experimental failure condition with more impulsive closure than did subjects with less achievement value. These results appear in the third column of Table 5. The correlation between valuation on achievement and impulsive closure scores in this condition was .42 ($p < .10$).

Space relations aptitude. Subjects with less of the kind of intellectual ability measured by the Space Relations Test also reacted more strongly to the experimental failure manipulation and produced more impulsive closure than subjects scoring higher on the Space Relations Test. These results appear in the third column of Table 6. Correlation between Space Relations Test scores and the index of impulsive closure was .41 ($p < .10$) within this experimental condition.

TABLE 4

AVERAGE SCORES ON COMBINED INDEX OF CLOSURE FOR SUBJECTS WITH HIGH AND LOW SELF-ESTEEM

Self-esteem	Control conditions		Threat condition	Control vs. Threat	
	Success (Ego involved)	Failure (Non-involved)	Failure (Ego involved)	Difference	p
Low	1.7	1.8	2.2	.4	.02
High	1.8	1.8	1.9	.1	*ns*
Difference	−.1	0	.3	.3	
p value of difference			$<.10$		

TABLE 5

AVERAGE SCORES ON COMBINED INDEX OF CLOSURE FOR SUBJECTS WITH HIGH AND LOW VALUATION OF ACHIEVEMENT

Valuation of achievement	Control conditions		Threat condition	Control vs. Threat	
	Success (Ego involved)	Failure (Non-involved)	Failure (Ego involved)	Difference	p
High	1.8	1.8	2.3	.5	.02
Low	1.7	1.9	2.0	.2	*ns*
Difference	.1	−.1	.3	.3	
p value of difference			.05		

TABLE 6

AVERAGE SCORES ON COMBINED INDEX OF CLOSURE FOR SUBJECTS WITH LOW AND HIGH PERFORMANCE ON SPACE RELATIONS TEST

Space relations performance	Control conditions		Threat condition	Control vs. Threat	
	Success (Ego involved)	Failure (Non-involved)	Failure (Ego involved)	Difference	p
Low	1.7	2.1	2.2	.3	.02
High	1.9	1.7	1.7	−.1	*ns*
Difference	−.2	.4	.5	.4	
p value of difference			$<.05$		

Interaction of Threat and Personality

Persons with low self-esteem, strong valuation on achievement, and low space relations

ability were by far the more vulnerable and more responsive to the experimental manipulations of failure. The effects of the experimental manipulation of threat occurred almost exclusively among these subjects, as seen in the top lines of Tables 4, 5, and 6. Subjects with high self-esteem, or low valuation on achievement, or high space relations ability tended to be immune to the experimental manipulation of failure.

DISCUSSION

Two different experimental manipulations have thus produced highly similar patterns of results. Low regard by peers (Dittes, 1959) and, in the present experiment, an experience of failure, even though anonymous, both resulted in the tendency for subjects to form positive and easy conclusions on varied types of conceptual tasks. The pair of results seems most easily accounted for by the assumptions (*a*) that low regard by peers and experience of failure both damage self-esteem, and (*b*) that cognitive closure has acquired the capacity to enhance self-esteem.

To seek out closure impulsively as a compensating source of self-esteem, then, becomes a defensive maneuver induced by the threat or damage experience. Finding certainty and clarity is reassuring, and when self-esteem is threatened, this defensive device may be called upon, if circumstances provide opportunity.

This interpretation is supported by the pattern of interaction between the experimental manipulations and the personality characteristics that people brought with them to the experiment. In each experiment, subjects in the lowest two-thirds of the self-esteem distribution were the ones who were vulnerable to the manipulation of threat. Persons scoring high on a self-esteem questionnaire presumably have other resources of self-esteem that they can readily recall or rehearse under conditions of threat or danger—this may be all that is represented by a high score on such a questionnaire. They are less dependent for self-esteem on the esteem of peers, on specific experiences of success, or on such defensive devices as impulsive closure. Good performance on the Space Relations Test seems best interpreted as providing another index of general abilities and other resources that can nourish a generally high self-esteem. Therefore it is consistent to find the identical pattern of interaction with the Space Relations Test as with the measure of self-esteem.

The index of valuation of achievement presumably represents a measure of the degree to which persons depend for their esteem on experiences of personal success and achievement of the kind involved in the experimental manipulation. Subjects scoring low on this measure are presumably more reliant on other sources for their self-esteem and less vulnerable to withdrawal of this source.

Possible Limitations of Sample

The sample of graduate students for this experiment was chosen as one which was expected to be particularly vulnerable to manipulations of failure, that is, a group with a generally high achievement value. Although there is no decisive independent evidence confirming this assumption, subjects achieved an average score of 4.2 on the six-point scale measuring valuation of achievement. All subjects rated themselves as above average on this scale; no subject achieved an average score of 3 or less. The generality of the results is therefore limited to subjects with correspondingly high achievement value. But this limitation is precisely what was predicted in the first instance and demonstrated in the interaction results in Table 5.

The argument might be advanced that the employment of impulsive closure as a defensive response to threat may be limited to such a sample of graduate students. This seems unlikely for the following reasons. The pressures in the academic environment generally tend to punish impulsive closure and to reward delayed closure of higher quality. Witness the stereotype of the egghead who can never make up his mind about anything. It seems more likely that the esteem enhancing properties of closure have been learned much earlier than college days and that the defensive use of impulsive closure takes place in spite of, rather than because of, the fact that these subjects are graduate students.

Low Correlation between Tasks

The fact that the two tasks used to measure impulsive closure have such a low correlation ($r = .10$) presents a serious problem, even though considering group averages they tend to show similar consequences of the experi-

mental procedures. The low correlation may be attributed to the fact that for each of the specific tasks, peculiar skills, interests, or background (in judging prose passages or in judging prospective roommates) are elicited and tend to override the experimental effects for a few individual subjects.

It is more likely, however, that closure has not been clearly enough defined conceptually so as to generate equivalent operations. The two operational measures may represent somewhat different variables. For example there may be an important distinction between what may be called perceptual closure and achievement closure—between the relatively passive recognition and assurance that there is order and system and clarity in the environment around one and the experience of producing for oneself clearcut cognitions and solutions out of obvious initial ambiguity. The roommate task may largely represent the former and the parable task the latter. The roommate task may be approached and completed without ever feeling that it is a "problem" to be resolved. The parable task tends to present a challenge to the subject to see whether he can discover and fashion clear meaning.

In this light, it is significant that only the parable task, not the roommate task, interacts with the measure of achievement valuation. There is a strong interaction between the measure of valuation of achievement and the experimental conditions for this task, as seen in Table 7. The roommate task shows no such trend. Table 7 also shows that this difference does not hold for the other important interactions. In the case of self-esteem and the Space Relations Test, both tasks tend to show similar trends, although in no case do they achieve the same level of significance as when they are combined, and the parable task tends to show slightly stronger trends.

The noninvolved failure control group performed inconsistently. They reported, on the direct measures, that they were disturbed and felt threatened by the manipulations but that this was not continually on their minds, as it was for the ego involved failure group (Table 1). On the parable task this control group showed as much impulsive behavior as the threat group. But on the roommate task they showed the least impulsive behavior (Table 3).

This greater variability of the group performance very likely simply reflects the greater variability of individuals within this group: on every measure obtained during the experimental session, it showed more variability than either of the others. This is presumably due to the greater ambiguity of the manipulations, in which some individuals would feel the threat of failure keenly and tend to ignore the noninvolving instructions, whereas other individuals would recall eagerly the noninvolving instructions and feel no threat.

There is a possibility, however, that the reassurance of the noninvolving instructions may have a delayed effect. The parable task was for all subjects the first one performed after the experimental manipulation, so that there

TABLE 7
CLOSURE SCORES ON PARABLE AND ROOMMATE TASKS

Personality measure		Parable task				Roommate task			
		Control		Threat	p Control vs. Threat	Control		Threat	p Control vs. Threat
		Success (Ego involved)	Failure (Nonin-volved)	Failure (Ego involved)		Success (Ego involved)	Failure (Nonin-volved)	Failure (Ego involved)	
Self-Esteem	Low	1.6	1.7	2.2	.10	1.9	1.9	2.1	<.10
	High	1.6	1.6	1.7		2.0	1.6	2.1	
p (Low vs. High)				.10					
Valuation of achievement	High	1.8	1.6	2.6	<.02	1.9	1.9	1.7	
	Low	1.5	2.0	1.8		2.0	1.8	2.0	
p (High vs. Low)				.07					
Space Relations Test	Low	1.8	2.3	2.2	<.15	1.6	2.1	2.0	<.15
	High	1.6	1.8	1.5		2.2	1.5	2.0	
p (Low vs. High)				<.05		<.02	<.02		

is a confounding of task and order. The experience of failure may have produced an immediate sense of disturbance and threat and the comforting thoughts that this test was valid only for stagehands and architects may have occurred only after an interval of time.

Notice, in Table 1, that these subjects report a strong sense of immediate threat (Lines 5 and 7) but that the threat did not stay on their minds throughout the experiment (Line 6). The parable task may have been performed before the noninvolving instructions took effect psychologically.

SUMMARY

Some persons impose meaning and pattern—closure—on ambiguous stimuli more readily than other persons. To account for such differences, the hypothesis is proposed that closure generally acquires a self-esteem enhancing reward value and that persons experiencing greater need to maintain or enhance self-esteem are therefore likely to impose more impulsive closure.

Self-esteem was manipulated by causing some subjects to experience failure on an ego involving task and others (two control groups) to experience failure on a noninvolving task or to experience success. Ego involving failure resulted in more impulsive closure as represented by tendencies (*a*) to find a positive meaning in an essentially incoherent prose passage rather than to acknowledge its incoherence, and (*b*) to base an impression of another person exclusively on more prominent traits and to ignore inconsistent traits.

Consistent with the self-esteem hypothesis, the success-failure manipulation did not affect closure equally for all subjects. Only the subjects low in self-esteem, low in a type of intellectual ability, or high in valuation of achievement were vulnerable to the success-failure manipulation and responded to it with the predicted differences in closure behavior.

These findings correspond exactly to those of a previous experiment in which self-esteem was manipulated by affecting the subjects' sense of acceptance by peers.

REFERENCES

ASCH, S. E. *Social psychology.* New York: Prentice-Hall, 1952.

DITTES, J. E. Effect of changes in self-esteem upon impulsiveness and deliberation in making judgments. *J. abnorm. soc. Psychol.*, 1959, **58**, 348–356.

FRENKEL-BRUNSWIK, ELSE. Intolerance of ambiguity as an emotional and perceptual personality variable. *J. Pers.*, 1949, **18**, 108–143.

GARDNER, R., HOLZMAN, P. S., KLEIN, G. S., LINTON, HARRIET, & SPENCE, D. P. Cognitive control, a study of individual consistencies in cognitive behavior. *Psychol. Issues*, 1959, **1**(4, Whole No. 4).

HOLZMAN, P. S. The relation of assimilation tendencies in visual, auditory, and kinesthetic time-error to cognitive attitudes of leveling and sharpening. *J. Pers.*, 1954, **22**, 375–394.

KLEIN, G. S. Cognitive control and motivation. In G. Lindzey (Ed.), *Assessment of human motives.* New York: Rinehart, 1958. Pp. 87–118.

MESSICK, S., & HILLS, J. R. Objective measurement of personality: Cautiousness and intolerance of ambiguity. *Educ. psychol. Measmt.*, 1960, in press.

MOFFITT, J. W., & STAGNER, R. Perceptual rigidity and closure as functions of anxiety. *J. abnorm. soc. Psychol.*, 1956, **52**, 354–357.

SHIPLEY, T. E. Threat to the self, the direction and breadth of attention, and the distance gradient. Unpublished doctoral dissertation, University of Pennsylvania, 1956.

(Received November 26, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 63, No. 3, 570-574

MOTIVE PATTERNS AND RISK PREFERENCES ASSOCIATED WITH ENTREPRENEURSHIP

HERBERT H. MEYER, WILLIAM B. WALKER
General Electric Company
AND GEORGE H. LITWIN
Harvard University

THE work of McClelland, Atkinson, and their collaborators (Atkinson, 1958; McClelland, Atkinson, Clark, & Lowell, 1953) in developing motive measures and in identifying correlates of these measures would seem to have some practical implications for improving predictions of certain aspects of individual performance in the industrial setting. This study was undertaken as the first phase of a research program designed to test a number of hypotheses regarding expected relationships between measures of motivation and certain attitudinal or behavioral variables in an industrial population.

The measurement of "need for achievement" (n Achievement), upon which a good deal of attention has been focused recently, would seem to be particularly pertinent in the competitive environment that characterizes many industrial roles. The fact that previous research has demonstrated a relationship between achievement motivation and risk taking behavior increases the possible contribution of this kind of research to the understanding and prediction of industrial role performance.

Atkinson (1957) presents a theoretical model showing that individual behavior in a situation of competitive achievement involving risk can be accounted for by the interaction of the "motive to achieve success" and the "motive to avoid failure." By assuming that motivation to take a particular course of action is a multiplicative function of (*a*) the strength of the motive, (*b*) the expectancy (subjective probability) of success, and (*c*) the value of the incentive, and by making the further assumption that in a situation involving risk the amount of satisfaction inherent in success is an inverse function of the strength of the expectancy (probability) of success, Atkinson demonstrates that a risk situation where chances of success are approximately 50-50 is most attractive to the person who is strongly motivated to achieve. The individual who is primarily motivated by the need to avoid failure, on the other hand, finds a risk situation where odds of success are approximately 50-50 most unattractive.

Several research studies provide empirical support for the hypotheses based on Atkinson's (1957) theoretical model. McClelland (1958) found in a study of school children that the tendency to prefer moderate rather than extreme risks on a ring toss game was significantly related to scores on a graphic measure of n Achievement. Atkinson and Litwin (1960) found in a study of college students that preference for intermediate level risks on a ring toss game was related to high n Achievement scores on the French Test of Insight, and preference for extreme risks was related to high scores on a Test Anxiety questionnaire. Atkinson, Bastian, Earl, and Litwin (1960), in a similar study of college students, found high n Achievement scores on the French Test of Insight were significantly related to the preference for intermediate risks on a shuffleboard game, and the preference for intermediate odds of winning over extreme odds of winning in a paper-and-pencil imaginary betting questionnaire. Litwin (1958) found in an experimental study of college students that high n Achievement and low test anxiety were related to preference for intermediate risks on a series of tasks including ring toss, penny pitch, a horse race game, and a series of pencil maze puzzles.

McClelland (1957) has argued that high achievement motivation should be associated with successful entrepreneurship. He has shown, for example, that the rate of economic growth of a number of different countries over the period from 1925 to 1950 was significantly correlated with the degree to which need for achievement themes were found in stories in widespread use in children's books in the respective cultures. He has also demonstrated that certain personality characteristics usually

associated with the entrepreneurial role, such as willingness to take risks and sense of responsibility, characterize the person who is high in achievement motivation (McClelland, 1961).

More recently, McClelland initiated a series of studies to compare the measured n Achievement of persons who have demonstrated success in jobs that are definitely entrepreneurial in character with that of persons in definitely nonentrepreneurial jobs. One of the reasons for undertaking the investigation reported here was to provide additional data for that series of studies.

METHOD

Selection of Groups

The study design required that measures of certain motives and risk preferences be obtained from persons in jobs with definite entrepreneurial characteristics and from a comparable sample of persons in jobs that are definitely nonentrepreneurial. An entrepreneurial job was defined as follows:

The entrepreneur should have more responsibility for *initiating* decisions, rather than merely having to make decisions when presented with problems.

He should also more often have *individual responsibility* for decisions and their effects—that is, without the review or assistance of a supervisor or a committee.

The entrepreneur should get *objective feedback* of accurate data indicating the success of his decisions, such as in sales volume, profitability, reduction in errors, complaints, etc., rather than just a general subjective evaluation of his success by the group or by superiors.

This type of feedback also means that the *job entails more risk* or challenge in that there is more chance of a serious wrong decision being observed.

The entrepreneurial and nonentrepreneurial groups were to be matched on the following variables: salary or status, age, years of education, and length of company service. If these factors were not held constant, differences that might be found on the experimental measures could be attributable to differences in age, education, or status.

Managers of shop operations in manufacturing plants were selected to represent the "entrepreneurial role." These managers had complete responsibility for administering specific parts of the manufacturing operation. In addition to the production employees in their groups, these managers had supporting personnel for such functions as scheduling, manufacturing engineering, and quality control reporting directly to them. Thus, in effect, each managed a fairly complete little business.

For each Manager in the sample a staff "Specialist" was chosen whose job seemed to have a minimum amount of entrepreneurial characteristics, and whose position level, age, educational background, and length of service was approximately the same as the respective Manager. Table 1 presents the means and ranges on these variables.

TABLE 1
CHARACTERISTICS OF THE MANAGERS AND SPECIALISTS PARTICIPATING IN THE STUDY

	Age		Education (Years)		Position level[a]		Years service	
	M	Range	*M*	Range	*M*	Range	*M*	Range
Managers ($N = 31$)	42.1	28–60	12.6	8–17	8.5	6–10	18.9	3–34
Specialists ($N = 31$)	42.7	27–59	13.2	10–17	7.7	5–10	18.1	3–42

[a] The position level refers to a company salary plan system. The difference shown would indicate that the salaries for the Managers were, on the average, about 6 or 7% higher than those of the Specialists.

On the average, the two groups were almost identical in age and years of service, but the Specialists were slightly higher in education and slightly lower in position level.[1]

Experimental Measures

Each participant completed a short Risk Preference Questionnaire and wrote brief stories to six thematic apperception pictures. The pictures were selected by McClelland specifically for a study of businessmen. They are described in Atkinson (1958) (Recommended Multiple Purpose Set A, Appendix III). The stories were scored for n Achievement, n Affiliation, and n Power by experienced scorers who had established rank-order scoring reliabilities of .90 or better prior to scoring the protocols for this study.[2]

The Risk Preference Questionnaire consisted merely of a list of pairs of betting choices with instructions to choose the preferred alternative. The odds ranged from one chance in six to six chances out of six of winning, with hypothetical rewards presented for each alternative at an amount related directly to the size of the odds. The test was scored by comparing the number of times the extreme risk options (1/6 and 5/6) were chosen to the number of times the intermediate risk options (2/6, 3/6, and 4/6) were chosen.

[1] The mean scores in Table 1 do not, of course, indicate the accuracy with which individual matches were made on the variables. Since there was a rather limited number of individuals available in any component for matching, compromises were necessary to get reasonably close matches on all four variables. The mean discrepancies in the variables for the individual matches were 4.68 years for age, .90 years for education, 1.06 in position level, and 5.42 in years of service. The distributions of discrepancies, however, were a little skewed, with most of the differences below the means. In each case a few very large discrepancies which were unavoidable inflated the averages.

[2] We wish to acknowledge the assistance of Ann Litwin who scored the protocols for n Achievement and n Power. The protocols for n Affiliation were scored by George Litwin.

Interviews

Each participant was subsequently interviewed to obtain an indication of his attitudes toward the entrepreneurial aspects of the Manager position. For the Specialists, the interviewer also attempted to determine to what degree the man's own job had entrepreneurial characteristics. Entrepreneurial characteristics considered were: the degree of objective risk the position entailed—that is, the chance to make a serious wrong decision, the result of which would be observable from objective indicators; responsibility for initiating decisions; and the objectivity of results measures fed back to the position incumbent. At the termination of the interview, the interviewer rated the degree to which that participant's attitudes were positive or negative in relation to these characteristics, and for the Specialists, he also rated the degree to which the man's job seemed to have these entrepreneurial characteristics. In addition, the interviewer also questioned each participant regarding his "level of aspiration"—the ultimate position or status for which he was striving in his career.

Management Ratings of "Achievement Drive"

Managers from two levels above the subjects in this study were asked to rank the participating Managers and Specialists on the degree to which they had demonstrated "Achievement Drive" on their jobs. Specifically, they were asked to indicate: "How strongly does the man seem to be motivated to do a good job? Does he identify with the job, work hard at it, take pride in doing the work, strive to improve, etc.?"

The ranks assigned to each man were translated into standard scores on a 10-point scale. The ranks given to any one man by the various raters were then averaged and rounded off to the nearest whole number to obtain the summary rating statistic used in the data analysis. Since the Managers and Specialists were in four different departments, each rater considered approximately eight participants in each of the two groups. Each participant was rated by at least two higher level managers, and some by more. The split-half estimate of reliability (half the judges vs. the other half on the same participants) was .77.

RESULTS

Motive and Risk Scores for Managers and Specialists

Table 2 presents differences found on the motive scores and the Risk Preference Questionnaire for the Managers and Specialists. Significant differences (by Wilcoxon matched-pairs signed-ranks test) were found between the groups on n Achievement and in the tendency to choose intermediate rather than extreme odds on the Risk Preference Questionnaire.[3]

TABLE 2

MEAN SCORES OF MANAGERS AND SPECIALISTS ON THE MOTIVE MEASURES AND THE RISK PREFERENCE QUESTIONNAIRE

	Managers	Specialists
Motive measures		
n Achievement	6.74	4.77**
n Affiliation	3.42	2.45 (*ns*)
n Power	6.39	6.45 (*ns*)
Risk Preference Questionnaire		
Percentage preferring intermediate odds:	52	32*

* Difference between matched pairs significant at .05 level.
** Difference between matched pairs significant at .025 level.

Motivation and Entrepreneurial Attitudes

None of the correlations between the interviewers' ratings of the Managers' and Specialists' attitudes toward the entrepreneurial aspects of the Manager position and motive scores were found to be statistically significant. A significant correlation was found, however, between a composite "entrepreneurial attitudes" score and observed "achievement drive" of the managers, as rated by their superiors (.42).[4]

For the Specialists' group only, correlations were also computed between the interviewers' ratings of the degree to which the Specialist's job seemed to have entrepreneurial characteristics and scores on n Achievement and the Risk Preference Questionnaire. These correlations (.01 and .12, respectively) were not significantly greater than zero.

Other Relationships

Intercorrelations of the predictor and control variables for the entire participant groups revealed few significant correlations. The n Affiliation measure was found to correlate negatively with age (−.35) and positively with education (.28), management ratings of achievement correlated negatively with age (−.33), and, as expected, age and education

[3] One-tailed tests were used in computing the significance of these differences since the direction of the differences was predicted. A two-tailed test was used in considering the differences in n Affiliation since a difference was not predicted. Although a difference was expected in n Power, it is obvious that none was found.

[4] Scores of participants on the motive measures and Risk Preference Questionnaire were withheld from the interviewers, and management ratings of the participants were not obtained at a plant until the interviewing had been completed, in order to insure that correlations between data gathered from different sources would not be contaminated.

were found to be negatively correlated (−.64). Probably of greater interest is the fact that n Achievement and risk preferences were not found to be significantly correlated with each other (−.01).

Analysis was also made of the relationships between the interviewers' ratings of each participant's "level of aspiration" and some of the other variables. Rated level of aspiration was found to be significantly correlated in a negative direction with age (−.48) and positively with education (.59), as might be expected. The correlation with n Achievement was also significant (.33), but if the effects of age and education are partialed out, this correlation is attenuated. Actually, level of aspiration ratings resulted in a skewed distribution, since almost half the participants (14 of the Managers and 14 of the Specialists) expressed no aspiration for a position beyond the level of the one they now held.

DISCUSSION

While the results of this exploratory study are encouraging, they still leave some questions unanswered. Two of the major hypotheses were verified, but other differences or relationships that had been expected were not found. Differences were found between the Manager group, selected to represent an entrepreneurial role, and the Specialists, whose jobs had a minimum of entrepreneurial characteristics, on n Achievement and preference for intermediate risks. These differences, however, were not at as high a level of significance as might be desirable. Furthermore, these variables did not correlate with attitudes of the subjects toward the entrepreneurial characteristics of a job. This negative finding may well be due, however, to the fact that these were relationships *within groups* where sample variances were too small to yield substantial correlations. The small samples and possible shortcomings in the measures could, of course, also have contributed to the lack of significant correlations.

Shortcomings in the two measures of achievement motivation could also account for the fact that they were found to be uncorrelated. Atkinson and his associates (Atkinson et al., 1960) in a study employing the same risk measure as was used here, also found that preference for intermediate risks was not significantly correlated with n Achievement. In the same study, however, they did find a significant correlation between n Achievement scores and preference for intermediate risks on a game of skill.

Risk taking behavior in a pure gambling situation may be a better indicator of fear of failure motivation, moreover, than of achievement or success oriented motivation. In this study the risk questionnaire was scored by combining preferences for extreme risks (1/6 or 5/6) and comparing this count with preferences for the intermediate level risks (2/6, 3/6, or 4/6). The Specialists, as a group, were found to prefer the extreme odds significantly more often than were the Managers. A close inspection of the data revealed, however, that the two groups actually did not differ in preference for the high risk (1/6) alternative. The difference in their scores was accounted for entirely by differences in their preference for the relatively low risk alternative (5/6), a finding that could indicate a fear of failure or "need for security" motivation.

The research evidence suggests that high fear of failure is not consistently associated with low n Achievement. It could very well be that these two factors are, at least in part, independent variables. Some research has suggested that high fear of failure may be associated with a moderate level of achievement motivation (McClelland et al., 1953). The man who is very low in achievement motivation may not have a strong enough need for success in a competitive situation to be greatly concerned about the consequences of failure.

If preference for low risk does indicate fear of failure motivation, and if this is characteristic of persons with a moderate level of need for achievement, we would expect the n Achievement scores for subjects in this study who showed preference for the 5/6 risk to cluster near the middle of the n Achievement distribution. A check of the data indicates a definite trend in this direction. The average n Achievement score for the men who preferred the 5/6 bet was a little lower, but not significantly different from the average score for the remainder of the sample. An analysis of the *distribution* of the scores, however, revealed that the variance of the n Achievement scores for the

men who preferred the 5/6 bet was significantly smaller than the variance of these scores for the remainder of the distribution.[5]

The fact that Managers' ratings of "achievement drive" did not correlate with either n Achievement or preference for intermediate risks is not surprising. Other researchers have reported similar findings. In one study, for example, self-ratings of achievement motivation were not found to be correlated with n Achievement (McClelland et al., 1953). In another study clinicians' ratings of need for achievement were not found to be related to n Achievement (McClelland et al., 1953), and in a third study a similar finding was reported in relation to teachers' ratings of the achievement drive of students (Winterbottom, 1958). The kind of achievement motivation that is tapped by a thematic apperceptive measure of n Achievement is not a personality characteristic which is easy to observe and evaluate.

SUMMARY

This exploratory study represents one of the first attempts in an industrial setting to determine the relationship between measured motive patterns of individuals and their adjustment to particular job roles. Persons in jobs with definite entrepreneurial characteristics were expected by hypothesis to score higher on a measure of "need for achievement" than persons whose jobs are nonentrepreneurial. Moreover, in a situation requiring decisions involving risk, entrepreneurs were expected to show preference for risks with odds of success near 50-50, since previous research had shown this type of risk behavior to be associated with achievement motivation.

Managers selected to represent the entrepreneurial role did score higher than Specialists of comparable age, education, and job level, whose jobs were judged to be nonentrepreneurial in nature, on a "projective" measure of achievement motivation, and showed greater preference for intermediate level risks on a Risk Preference Questionnaire. However, an attempt to establish more positively the fact that these differences could be attributed to the entrepreneurial characteristics of the roles of the individuals in the two groups was unsuccessful.

[5] F test significant at .01 level.

REFERENCES

ATKINSON, J. W. Motivational determinants of risk-taking behavior. *Psychol. Rev.*, 1957, **64**, 359–372.

ATKINSON, J. W. (Ed.) *Motives in fantasy, action, and society*. Princeton: Van Nostrand, 1958. Ch. 22.

ATKINSON, J. W., BASTIAN, J. R., EARL, R. W., & LITWIN, G. H. The achievement motive, goal-setting, and probability preferences. *J. abnorm. soc. Psychol.*, 1960, **60**, 27–36.

ATKINSON, J. W., & LITWIN, G. H. Achievement motive and test anxiety conceived as motive to approach success and motive to avoid failure. *J. abnorm. soc. Psychol.*, 1960, **60**, 52–63.

LITWIN, G. H. Motives and expectancy as determinants of preference for degrees of risk. Unpublished honors thesis, University of Michigan, 1958.

MCCLELLAND, D. C. Community development and the nature of human motivation: Some implications of recent research. Paper prepared for Conference on Community Development sponsored by Center for International Studies, Endicott House, Dec. 13–15, 1957.

MCCLELLAND, D. C. Risk taking in children with high and low need for achievement. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society*. Princeton: Van Nostrand, 1958. Pp. 306–321.

MCCLELLAND, D. C. *The achieving society*. Princeton: Van Nostrand, 1961. Ch. 6.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.

WINTERBOTTOM, M. R. The relation of need for achievement to learning experiences in independence and mastery. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society*. Princeton: Van Nostrand, 1958. Pp. 453–478.

(Received November 28, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 63, No. 3, 575-582

TRANSMISSION OF AGGRESSION THROUGH IMITATION OF AGGRESSIVE MODELS[1]

ALBERT BANDURA, DOROTHEA ROSS, AND SHEILA A. ROSS[2]
Stanford University

A previous study, designed to account for the phenomenon of identification in terms of incidental learning, demonstrated that children readily imitated behavior exhibited by an adult model in the presence of the model (Bandura & Huston, 1961). A series of experiments by Blake (1958) and others (Grosser, Polansky, & Lippitt, 1951; Rosenblith, 1959; Schachter & Hall, 1952) have likewise shown that mere observation of responses of a model has a facilitating effect on subjects' reactions in the immediate social influence setting.

While these studies provide convincing evidence for the influence and control exerted on others by the behavior of a model, a more crucial test of imitative learning involves the generalization of imitative response patterns to new settings in which the model is absent.

In the experiment reported in this paper children were exposed to aggressive and nonaggressive adult models and were then tested for amount of imitative learning in a new situation in the absence of the model. According to the prediction, subjects exposed to aggressive models would reproduce aggressive acts resembling those of their models and would differ in this respect both from subjects who observed nonaggressive models and from those who had no prior exposure to any models. This hypothesis assumed that subjects had learned imitative habits as a result of prior reinforcement, and these tendencies would generalize to some extent to adult experimenters (Miller & Dollard, 1941).

It was further predicted that observation of subdued nonaggressive models would have a generalized inhibiting effect on the subjects' subsequent behavior, and this effect would be reflected in a difference between the nonaggressive and the control groups, with subjects in the latter group displaying significantly more aggression.

Hypotheses were also advanced concerning the influence of the sex of model and sex of subjects on imitation. Fauls and Smith (1956) have shown that preschool children perceive their parents as having distinct preferences regarding sex appropriate modes of behavior for their children. Their findings, as well as informal observation, suggest that parents reward imitation of sex appropriate behavior and discourage or punish sex inappropriate imitative responses, e.g., a male child is unlikely to receive much reward for performing female appropriate activities, such as cooking, or for adopting other aspects of the maternal role, but these same behaviors are typically welcomed if performed by females. As a result of differing reinforcement histories, tendencies to imitate male and female models thus acquire differential habit strength. One would expect, on this basis, subjects to imitate the behavior of a same-sex model to a greater degree than a model of the opposite sex.

Since aggression, however, is a highly masculine-typed behavior, boys should be more predisposed than girls toward imitating aggression, the difference being most marked for subjects exposed to the male aggressive model.

METHOD

Subjects

The subjects were 36 boys and 36 girls enrolled in the Stanford University Nursery School. They ranged in age from 37 to 69 months, with a mean age of 52 months.

Two adults, a male and a female, served in the role of model, and one female experimenter conducted the study for all 72 children.

Experimental Design

Subjects were divided into eight experimental groups of six subjects each and a control group consisting of 24 subjects. Half the experimental subjects were exposed to aggressive models and half were exposed to models that were subdued and nonaggressive in their behavior. These groups were further subdivided into male and female subjects. Half the subjects in the aggressive and nonaggressive conditions observed

[1] This investigation was supported by Research Grant M-4398 from the National Institute of Health, United States Public Health Service.

[2] The authors wish to express their appreciation to Edith Dowley, Director, and Patricia Rowe, Head Teacher, Stanford University Nursery School for their assistance throughout this study.

same-sex models, while the remaining subjects in each group viewed models of the opposite sex. The control group had no prior exposure to the adult models and was tested only in the generalization situation.

It seemed reasonable to expect that the subjects' level of aggressiveness would be positively related to the readiness with which they imitated aggressive modes of behavior. Therefore, in order to increase the precision of treatment comparisons, subjects in the experimental and control groups were matched individually on the basis of ratings of their aggressive behavior in social interactions in the nursery school.

The subjects were rated on four five-point rating scales by the experimenter and a nursery school teacher, both of whom were well acquainted with the children. These scales measured the extent to which subjects displayed physical aggression, verbal aggression, aggression toward inanimate objects, and aggressive inhibition. The latter scale, which dealt with the subjects' tendency to inhibit aggressive reactions in the face of high instigation, provided a measure of aggression anxiety.

Fifty-one subjects were rated independently by both judges so as to permit an assessment of interrater agreement. The reliability of the composite aggression score, estimated by means of the Pearson product-moment correlation, was .89.

The composite score was obtained by summing the ratings on the four aggression scales; on the basis of these scores, subjects were arranged in triplets and assigned at random to one of two treatment conditions or to the control group.

Experimental Conditions

In the first step in the procedure subjects were brought individually by the experimenter to the experimental room and the model who was in the hallway outside the room, was invited by the experimenter to come and join in the game. The experimenter then escorted the subject to one corner of the room, which was structured as the subject's play area. After seating the child at a small table, the experimenter demonstrated how the subject could design pictures with potato prints and picture stickers provided. The potato prints included a variety of geometrical forms; the stickers were attractive multicolor pictures of animals, flowers, and western figures to be pasted on a pastoral scene. These activities were selected since they had been established, by previous studies in the nursery school, as having high interest value for the children.

After having settled the subject in his corner, the experimenter escorted the model to the opposite corner of the room which contained a small table and chair, a tinker toy set, a mallet, and a 5-foot inflated Bobo doll. The experimenter explained that these were the materials provided for the model to play with and, after the model was seated, the experimenter left the experimental room.

With subjects in the *nonaggressive condition*, the model assembled the tinker toys in a quiet subdued manner totally ignoring the Bobo doll.

In contrast, with subjects in the *aggressive condition*, the model began by assembling the tinker toys but after approximately a minute had elapsed, the model turned to the Bobo doll and spent the remainder of the period aggressing toward it.

Imitative learning can be clearly demonstrated if a model performs sufficiently novel patterns of responses which are unlikely to occur independently of the observation of the behavior of a model and if a subject reproduces these behaviors in substantially identical form. For this reason, in addition to punching the Bobo doll, a response that is likely to be performed by children independently of a demonstration, the model exhibited distinctive aggressive acts which were to be scored as imitative responses. The model laid Bobo on its side, sat on it and punched it repeatedly in the nose. The model then raised the Bobo doll, picked up the mallet and struck the doll on the head. Following the mallet aggression, the model tossed the doll up in the air aggressively and kicked it about the room. This sequence of physically aggressive acts was repeated approximately three times, interspersed with verbally aggressive responses such as, "Sock him in the nose . . . ," "Hit him down . . . ," "Throw him in the air . . . ," "Kick him . . . ," "Pow . . . ," and two nonaggressive comments, "He keeps coming back for more" and "He sure is a tough fella."

Thus in the exposure situation, subjects were provided with a diverting task which occupied their attention while at the same time insured observation of the model's behavior in the absence of any instructions to observe or to learn the responses in question. Since subjects could not perform the model's aggressive behavior, any learning that occurred was purely on an observational or covert basis.

At the end of 10 minutes, the experimenter entered the room, informed the subject that he would now go to another game room, and bid the model goodbye.

Aggression Arousal

Subjects were tested for the amount of imitative learning in a different experimental room that was set off from the main nursery school building. The two experimental situations were thus clearly differentiated; in fact, many subjects were under the impression that they were no longer on the nursery school grounds.

Prior to the test for imitation, however, all subjects, experimental and control, were subjected to mild aggression arousal to insure that they were under some degree of instigation to aggression. The arousal experience was included for two main reasons. In the first place, observation of aggressive behavior exhibited by others tends to reduce the probability of aggression on the part of the observer (Rosenbaum & deCharms, 1960). Consequently, subjects in the aggressive condition, in relation both to the nonaggressive and control groups, would be under weaker instigation following exposure to the models. Second, if subjects in the nonaggressive condition expressed little aggression in the face of appropriate instigation, the presence of an inhibitory process would seem to be indicated.

Following the exposure experience, therefore, the experimenter brought the subject to an anteroom that contained these relatively attractive toys: a fire engine, a locomotive, a jet fighter plane, a cable car, a colorful spinning top, and a doll set complete with wardrobe, doll carriage, and baby crib. The experi-

menter explained that the toys were for the subject to play with but, as soon as the subject became sufficiently involved with the play material (usually in about 2 minutes), the experimenter remarked that these were her very best toys, that she did not let just anyone play with them, and that she had decided to reserve these toys for the other children. However, the subject could play with any of the toys that were in the next room. The experimenter and the subject then entered the adjoining experimental room.

It was necessary for the experimenter to remain in the room during the experimental session; otherwise a number of the children would either refuse to remain alone or would leave before the termination of the session. However, in order to minimize any influence her presence might have on the subject's behavior, the experimenter remained as inconspicuous as possible by busying herself with paper work at a desk in the far corner of the room and avoiding any interaction with the child.

Test for Delayed Imitation

The experimental room contained a variety of toys including some that could be used in imitative or nonimitative aggression, and others that tended to elicit predominantly nonaggressive forms of behavior. The aggressive toys included a 3-foot Bobo doll, a mallet and peg board, two dart guns, and a tether ball with a face painted on it which hung from the ceiling. The nonaggressive toys, on the other hand, included a tea set, crayons and coloring paper, a ball, two dolls, three bears, cars and trucks, and plastic farm animals.

In order to eliminate any variation in behavior due to mere placement of the toys in the room, the play material was arranged in a fixed order for each of the sessions.

The subject spent 20 minutes in this experimental room during which time his behavior was rated in terms of predetermined response categories by judges who observed the session through a one-way mirror in an adjoining observation room. The 20-minute session was divided into 5-second intervals by means of an electric interval timer, thus yielding a total number of 240 response units for each subject.

The male model scored the experimental sessions for all 72 children. Except for the cases in which he served as model, he did not have knowledge of the subjects' group assignments. In order to provide an estimate of interscorer agreement, the performances of half the subjects were also scored independently by a second observer. Thus one or the other of the two observers usually had no knowledge of the conditions to which the subjects were assigned. Since, however, all but two of the subjects in the aggressive condition performed the models' novel aggressive responses while subjects in the other conditions only rarely exhibited such reactions, subjects who were exposed to the aggressive models could be readily identified through their distinctive behavior.

The responses scored involved highly specific concrete classes of behavior and yielded high interscorer reliabilities, the product-moment coefficients being in the .90s.

Response Measures

Three measures of imitation were obtained:

Imitation of physical aggression: This category included acts of striking the Bobo doll with the mallet, sitting on the doll and punching it in the nose, kicking the doll, and tossing it in the air.

Imitative verbal aggression: Subject repeats the phrases, "Sock him," "Hit him down," "Kick him," "Throw him in the air," or "Pow."

Imitative nonaggressive verbal responses: Subject repeats, "He keeps coming back for more," or "He sure is a tough fella."

During the pretest, a number of the subjects imitated the essential components of the model's behavior but did not perform the complete act, or they directed the imitative aggressive response to some object other than the Bobo doll. Two responses of this type were therefore scored and were interpreted as partially imitative behavior.

Mallet aggression: Subject strikes objects other than the Bobo doll aggressively with the mallet.

Sits on Bobo doll: Subject lays the Bobo doll on its side and sits on it, but does not aggress toward it.

The following additional nonimitative aggressive responses were scored:

Punches Bobo doll: Subject strikes, slaps, or pushes the doll aggressively.

Nonimitative physical and verbal aggression: This category included physically aggressive acts directed toward objects other than the Bobo doll and any hostile remarks except for those in the verbal imitation category; e.g., "Shoot the Bobo," "Cut him," "Stupid ball," "Knock over people," "Horses fighting, biting"

Aggressive gun play: Subject shoots darts or aims the guns and fires imaginary shots at objects in the room.

Ratings were also made of the number of behavior units in which subjects played nonaggressively or sat quietly and did not play with any of the material at all.

RESULTS

Complete Imitation of Models' Behavior

Subjects in the aggression condition reproduced a good deal of physical and verbal aggressive behavior resembling that of the models, and their mean scores differed markedly from those of subjects in the nonaggressive and control groups who exhibited virtually no imitative aggression (see Table 1).

Since there were only a few scores for subjects in the nonaggressive and control conditions (approximately 70% of the subjects had zero scores), and the assumption of homogeneity of variance could not be made, the Friedman two-way analysis of variance by ranks was employed to test the significance of the obtained differences.

The prediction that exposure of subjects to aggressive models increases the probability

of aggressive behavior is clearly confirmed (see Table 2). The main effect of treatment conditions is highly significant both for physical and verbal imitative aggression. Comparison of pairs of scores by the sign test shows that the obtained over-all differences were due almost entirely to the aggression displayed by subjects who had been exposed to the aggressive models. Their scores were significantly higher than those of either the nonaggressive or control groups, which did not differ from each other (Table 2).

Imitation was not confined to the model's aggressive responses. Approximately one-third of the subjects in the aggressive condition also repeated the model's nonaggressive verbal responses while none of the subjects in either the nonaggressive or control groups made such remarks. This difference, tested by means of the Cochran Q test, was significant well beyond the .001 level (Table 2).

TABLE 1
MEAN AGGRESSION SCORES FOR EXPERIMENTAL AND CONTROL SUBJECTS

Response category	Experimental groups				Control groups
	Aggressive		Nonaggressive		
	F Model	M Model	F Model	M Model	
Imitative physical aggression					
Female subjects	5.5	7.2	2.5	0.0	1.2
Male subjects	12.4	25.8	0.2	1.5	2.0
Imitative verbal aggression					
Female subjects	13.7	2.0	0.3	0.0	0.7
Male subjects	4.3	12.7	1.1	0.0	1.7
Mallet aggression					
Female subjects	17.2	18.7	0.5	0.5	13.1
Male subjects	15.5	28.8	18.7	6.7	13.5
Punches Bobo doll					
Female subjects	6.3	16.5	5.8	4.3	11.7
Male subjects	18.9	11.9	15.6	14.8	15.7
Nonimitative aggression					
Female subjects	21.3	8.4	7.2	1.4	6.1
Male subjects	16.2	36.7	26.1	22.3	24.6
Aggressive gun play					
Female subjects	1.8	4.5	2.6	2.5	3.7
Male subjects	7.3	15.9	8.9	16.7	14.3

Partial Imitation of Models' Behavior

Differences in the predicted direction were also obtained on the two measures of partial imitation.

Analysis of variance of scores based on the subjects' use of the mallet aggressively toward objects other than the Bobo doll reveals that treatment conditions are a statistically significant source of variation (Table 2). In addition, individual sign tests show that both the aggressive and the control groups, relative to subjects in the nonaggressive condition, produced significantly more mallet aggression, the difference being particularly marked with regard to female subjects. Girls who observed nonaggressive models performed a mean number of 0.5 mallet aggression responses as compared to mean values of 18.0 and 13.1 for girls in the aggressive and control groups, respectively.

Although subjects who observed aggressive models performed more mallet aggression (M = 20.0) than their controls (M = 13.3), the difference was not statistically significant.

TABLE 2
SIGNIFICANCE OF THE DIFFERENCES BETWEEN EXPERIMENTAL AND CONTROL GROUPS IN THE EXPRESSION OF AGGRESSION

Response category	χ^2_r	Q	p	Comparison of pairs of treatment conditions		
				Aggressive vs. Nonaggressive p	Aggressive vs. Control p	Nonaggressive vs. Control p
Imitative responses						
Physical aggression	27.17		<.001	<.001	<.001	.09
Verbal aggression	9.17		<.02	.004	.048	.09
Nonaggressive verbal responses		17.50	<.001	.004	.004	*ns*
Partial imitation						
Mallet aggression	11.06		<.01	.026	*ns*	.005
Sits on Bobo		13.44	<.01	.018	.059	*ns*
Nonimitative aggression						
Punches Bobo doll	2.87		*ns*			
Physical and verbal	8.96		<.02	.026	*ns*	*ns*
Aggressive gun play	2.75		*ns*			

With respect to the partially imitative response of sitting on the Bobo doll, the over-all group differences were significant beyond the .01 level (Table 2). Comparison of pairs of scores by the sign test procedure reveals that subjects in the aggressive group reproduced this aspect of the models' behavior to a greater extent than did the nonaggressive ($p = .018$) or the control ($p = .059$) subjects. The latter two groups, on the other hand, did not differ from each other.

Nonimitative Aggression

Analyses of variance of the remaining aggression measures (Table 2) show that treatment conditions did not influence the extent to which subjects engaged in aggressive gun play or punched the Bobo doll. The effect of conditions is highly significant ($\chi^2_r = 8.96$, $p < .02$), however, in the case of the subjects' expression of nonimitative physical and verbal aggression. Further comparison of treatment pairs reveals that the main source of the over-all difference was the aggressive and nonaggressive groups which differed significantly from each other (Table 2), with subjects exposed to the aggressive models displaying the greater amount of aggression.

Influence of Sex of Model and Sex of Subjects on Imitation

The hypothesis that boys are more prone than girls to imitate aggression exhibited by a model was only partially confirmed. t tests computed for subjects in the aggressive condition reveal that boys reproduced more imitative physical aggression than girls ($t = 2.50$, $p < .01$). The groups do not differ, however, in their imitation of verbal aggression.

The use of nonparametric tests, necessitated by the extremely skewed distributions of scores for subjects in the nonaggressive and control conditions, preclude an over-all test of the influence of sex of model per se, and of the various interactions between the main effects. Inspection of the means presented in Table 1 for subjects in the aggression condition, however, clearly suggests the possibility of a Sex × Model interaction. This interaction effect is much more consistent and pronounced for the male model than for the female model. Male subjects, for example, exhibited more physical ($t = 2.07$, $p < .05$) and verbal imitative aggression ($t = 2.51$, $p < .05$), more nonimitative aggression ($t = 3.15$, $p < .025$), and engaged in significantly more aggressive gun play ($t = 2.12$, $p < .05$) following exposure to the aggressive male model than the female subjects. In contrast, girls exposed to the female model performed considerably more imitative verbal aggression and more nonimitative aggression than did the boys (Table 1). The variances, however, were equally large and with only a small N in each cell the mean differences did not reach statistical significance.

Data for the nonaggressive and control subjects provide additional suggestive evidence that the behavior of the male model exerted a greater influence than the female model on the subjects' behavior in the generalization situation.

It will be recalled that, except for the greater amount of mallet aggression exhibited by the control subjects, no significant differences were obtained between the nonaggressive and control groups. The data indicate, however, that the absence of significant differences between these two groups was due primarily to the fact that subjects exposed to the nonaggressive female model did not differ from the controls on any of the measures of aggression. With respect to the male model, on the other hand, the differences between the groups are striking. Comparison of the sets of scores by means of the sign test reveals that, in relation to the control group, subjects exposed to the nonaggressive male model performed significantly less imitative physical aggression ($p = .06$), less imitative verbal aggression ($p = .002$), less mallet aggression ($p = .003$), less nonimitative physical and verbal aggression ($p = .03$), and they were less inclined to punch the Bobo doll ($p = .07$).

While the comparison of subgroups, when some of the over-all tests do not reach statistical significance, is likely to capitalize on chance differences, nevertheless the consistency of the findings adds support to the interpretation in terms of influence by the model.

Nonaggressive Behavior

With the exception of expected sex differences, Lindquist (1956) Type III analyses of variance of the nonaggressive response scores yielded few significant differences.

Female subjects spent more time than boys

playing with dolls ($p < .001$), with the tea set ($p < .001$), and coloring ($p < .05$). The boys, on the other hand, devoted significantly more time than the girls to exploratory play with the guns ($p < .01$). No sex differences were found in respect to the subjects use of the other stimulus objects, i.e., farm animals, cars, or tether ball.

Treatment conditions did produce significant differences on two measures of nonaggressive behavior that are worth mentioning. Subjects in the nonaggressive condition engaged in significantly more nonaggressive play with dolls than either subjects in the aggressive group ($t = 2.67$, $p < .02$), or in the control group ($t = 2.57$, $p < .02$).

Even more noteworthy is the finding that subjects who observed nonaggressive models spent more than twice as much time as subjects in aggressive condition ($t = 3.07$, $p < .01$) in simply sitting quietly without handling any of the play material.

DISCUSSION

Much current research on social learning is focused on the shaping of new behavior through rewarding and punishing consequences. Unless responses are emitted, however, they cannot be influenced. The results of this study provide strong evidence that observation of cues produced by the behavior of others is one effective means of eliciting certain forms of responses for which the original probability is very low or zero. Indeed, social imitation may hasten or short-cut the acquisition of new behaviors without the necessity of reinforcing successive approximations as suggested by Skinner (1953).

Thus subjects given an opportunity to observe aggressive models later reproduced a good deal of physical and verbal aggression (as well as nonaggressive responses) substantially identical with that of the model. In contrast, subjects who were exposed to nonaggressive models and those who had no previous exposure to any models only rarely performed such responses.

To the extent that observation of adult models displaying aggression communicates permissiveness for aggressive behavior, such exposure may serve to weaken inhibitory responses and thereby to increase the probability of aggressive reactions to subsequent frustrations. The fact, however, that subjects expressed their aggression in ways that clearly resembled the novel patterns exhibited by the models provides striking evidence for the occurrence of learning by imitation.

In the procedure employed by Miller and Dollard (1941) for establishing imitative behavior, adult or peer models performed discrimination responses following which they were consistently rewarded, and the subjects were similarly reinforced whenever they matched the leaders' choice responses. While these experiments have been widely accepted as demonstrations of learning by means of imitation, in fact, they simply involve a special case of discrimination learning in which the behavior of others serves as discriminative stimuli for responses that are already part of the subject's repertoire. Auditory or visual environmental cues could easily have been substituted for the social stimuli to facilitate the discrimination learning. In contrast, the process of imitation studied in the present experiment differed in several important respects from the one investigated by Miller and Dollard in that subjects learned to combine fractional responses into relatively complex novel patterns solely by observing the performance of social models without any opportunity to perform the models' behavior in the exposure setting, and without any reinforcers delivered either to the models or to the observers.

An adequate theory of the mechanisms underlying imitative learning is lacking. The explanations that have been offered (Logan, Olmsted, Rosner, Schwartz, & Stevens, 1955; Maccoby, 1959) assume that the imitator performs the model's responses covertly. If it can be assumed additionally that rewards and punishments are self-administered in conjunction with the covert responses, the process of imitative learning could be accounted for in terms of the same principles that govern instrumental trial-and-error learning. In the early stages of the developmental process, however, the range of component responses in the organism's repertoire is probably increased through a process of classical conditioning (Bandura & Huston, 1961; Mowrer, 1950).

The data provide some evidence that the male model influenced the subjects' behavior

outside the exposure setting to a greater extent than was true for the female model. In the analyses of the Sex × Model interactions, for example, only the comparisons involving the male model yielded significant differences. Similarly, subjects exposed to the nonaggressive male model performed less aggressive behavior than the controls, whereas comparisons involving the female model were consistently nonsignificant.

In a study of learning by imitation, Rosenblith (1959) has likewise found male experimenters more effective than females in influencing childrens' behavior. Rosenblith advanced the tentative explanation that the school setting may involve some social deprivation in respect to adult males which, in turn, enhances the male's reward value.

The trends in the data yielded by the present study suggest an alternative explanation. In the case of a highly masculine-typed behavior such as physical aggression, there is a tendency for both male and female subjects to imitate the male model to a greater degree than the female model. On the other hand, in the case of verbal aggression, which is less clearly sex linked, the greatest amount of imitation occurs in relation to the same-sex model. These trends together with the finding that boys in relation to girls are in general more imitative of physical aggression but do not differ in imitation of verbal aggression, suggest that subjects may be differentially affected by the sex of the model but that predictions must take into account the degree to which the behavior in question is sex-typed.

The preceding discussion has assumed that maleness-femaleness rather than some other personal characteristics of the particular models involved, is the significant variable—an assumption that cannot be tested directly with the data at hand. It was clearly evident, however, particularly from boys' spontaneous remarks about the display of aggression by the female model, that some subjects at least were responding in terms of a sex discrimination and their prior learning about what is sex appropriate behavior (e.g., "Who is that lady. That's not the way for a lady to behave. Ladies are supposed to act like ladies. . . ." "You should have seen what that girl did in there. She was just acting like a man. I never saw a girl act like that before. She was punching and fighting but no swearing."). Aggression by the male model, on the other hand, was more likely to be seen as appropriate and approved by both the boys ("Al's a good socker, he beat up Bobo. I want to sock like Al.") and the girls ("That man is a strong fighter, he punched and punched and he could hit Bobo right down to the floor and if Bobo got up he said, 'Punch your nose.' He's a good fighter like Daddy.").

The finding that subjects exposed to the quiet models were more inhibited and unresponsive than subjects in the aggressive condition, together with the obtained difference on the aggression measures, suggests that exposure to inhibited models not only decreases the probability of occurrence of aggressive behavior but also generally restricts the range of behavior emitted by the subjects.

"Identification with aggressor" (Freud, 1946) or "defensive identification" (Mowrer, 1950), whereby a person presumably transforms himself from object to agent of aggression by adopting the attributes of an aggressive threatening model so as to allay anxiety, is widely accepted as an explanation of the imitative learning of aggression.

The development of aggressive modes of response by children of aggressively punitive adults, however, may simply reflect object displacement without involving any such mechanism of defensive identification. In studies of child training antecedents of aggressively antisocial adolescents (Bandura & Walters, 1959) and of young hyperaggressive boys (Bandura, 1960), the parents were found to be nonpermissive and punitive of aggression directed toward themselves. On the other hand, they actively encouraged and reinforced their sons' aggression toward persons outside the home. This pattern of differential reinforcement of aggressive behavior served to inhibit the boys' aggression toward the original instigators and fostered the displacement of aggression toward objects and situations eliciting much weaker inhibitory responses.

Moreover, the findings from an earlier study (Bandura & Huston, 1961), in which children imitated to an equal degree aggression exhibited by a nurturant and a nonnurturant model, together with the results

of the present experiment in which subjects readily imitated aggressive models who were more or less neutral figures suggest that mere observation of aggression, regardless of the quality of the model-subject relationship, is a sufficient condition for producing imitative aggression in children. A comparative study of the subjects' imitation of aggressive models who are feared, who are linked and esteemed, or who are essentially neutral figures would throw some light on whether or not a more parsimonious theory than the one involved in "identification with the aggressor" can explain the modeling process.

SUMMARY

Twenty-four preschool children were assigned to each of three conditions. One experimental group observed aggressive adult models; a second observed inhibited nonaggressive models; while subjects in a control group had no prior exposure to the models. Half the subjects in the experimental conditions observed same-sex models and half viewed models of the opposite sex. Subjects were then tested for the amount of imitative as well as nonimitative aggression performed in a new situation in the absence of the models.

Comparison of the subjects' behavior in the generalization situation revealed that subjects exposed to aggressive models reproduced a good deal of aggression resembling that of the models, and that their mean scores differed markedly from those of subjects in the nonaggressive and control groups. Subjects in the aggressive condition also exhibited significantly more partially imitative and nonimitative aggressive behavior and were generally less inhibited in their behavior than subjects in the nonaggressive condition.

Imitation was found to be differentially influenced by the sex of the model with boys showing more aggression than girls following exposure to the male model, the difference being particularly marked on highly masculine-typed behavior.

Subjects who observed the nonaggressive models, especially the subdued male model, were generally less aggressive than their controls.

The implications of the findings based on this experiment and related studies for the psychoanalytic theory of identification with the aggressor were discussed.

REFERENCES

BANDURA, A. Relationship of family patterns to child behavior disorders. Progress Report, 1960, Stanford University, Project No. M-1734, United States Public Health Service.

BANDURA, A., & HUSTON, ALETHA C. Identification as a process of incidental learning. *J. abnorm. soc. Psychol.*, 1961, **63**, 311–318.

BANDURA, A., & WALTERS, R. H. *Adolescent aggression.* New York: Ronald, 1959.

BLAKE, R. R. The other person in the situation. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior.* Stanford, Calif: Stanford Univer. Press, 1958. Pp. 229–242.

FAULS, LYDIA B., & SMITH, W. D. Sex-role learning of five-year olds. *J. genet. Psychol.*, 1956, **89**, 105–117.

FREUD, ANNA. *The ego and the mechanisms of defense.* New York: International Univer. Press, 1946.

GROSSER, D., POLANSKY, N., & LIPPITT, R. A laboratory study of behavior contagion. *Hum. Relat.*, 1951, **4**, 115–142.

LINDQUIST, E. F. *Design and analysis of experiments.* Boston: Houghton Mifflin, 1956.

LOGAN, F., OLMSTED, O. L., ROSNER, B. S., SCHWARTZ, R. D., & STEVENS, C. M. *Behavior theory and social science.* New Haven: Yale Univer. Press, 1955.

MACCOBY, ELEANOR E. Role-taking in childhood and its consequences for social learning. *Child Develpm.*, 1959, **30**, 239–252.

MILLER, N. E., & DOLLARD, J. *Social learning and imitation.* New Haven: Yale Univer. Press, 1941.

MOWRER, O. H. (Ed.) Identification: A link between learning theory and psychotherapy. In, *Learning theory and personality dynamics.* New York: Ronald, 1950. Pp. 69–94.

ROSENBAUM, M. E., & DECHARMS, R. Direct and vicarious reduction of hostility. *J. abnorm. soc. Psychol.*, 1960, **60**, 105–111.

ROSENBLITH, JUDY F. Learning by imitation in kindergarten children. *Child Develpm.*, 1959, **30**, 69–80.

SCHACHTER, S., & HALL, R. Group-derived restraints and audience persuasion. *Hum. Relat.*, 1952, **5**, 397–406.

SKINNER, B. F. *Science and human behavior.* New York: Macmillan, 1953.

(Received December 2, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 63, No. 3, 583-587

THE CONDITIONING AND GENERALIZATION OF CHANGES IN SOCIAL PERCEPTUAL DISPOSITIONS[1]

LEON H. LEVY[2]
Indiana University

IMPRESSIONS formed of others are only partially determined by their stimulus qualities. In addition, the judge's overall attitude toward the person being judged (Thorndike, 1920), his "theory" of how traits are related to each other (Hays, 1958; Newcomb, 1931), his emotional state (Feshbach & Singer, 1957; Murray, 1933), the underlying structure of perception itself (Asch, 1946; Levy & Dugan, 1960; Wishner, 1960), and the context in which the stimulus person is presented for judgment (Levy, 1960, 1961) have all been shown to enter into the process of impression formation. Yet to be investigated, however, is the role played by learning in impression formation, and it is to this problem that the present investigation is addressed.

Using what was called a "constant error approach" Levy and Dugan (1960) obtained correlations between trait judgments which, by the nature of the method used, they claimed could not be attributed to either the operation of the halo effect or the judges' implicit theories of personality, but which must instead represent certain underlying dimensions of social perception. In essence, the constant error approach involves obtaining scores for each judge representing the frequency with which he makes a particular judgment on each of several bipolar dimensions where each judgment involves a different facial photograph, and then correlating these scores over judges. Considering the scores as representing tendencies to make particular judgments, or perceptual dispositions, the correlations between them may be viewed as reflecting the extent to which these dispositions are related to each other, and factors derived from a matrix of such correlations may be conceived of as representing certain basic dimensions of social perception or judgment.

Two questions immediately suggest themselves for investigation in regard to Levy and Dugan's (1960) findings. The first concerns the extent to which perceptual dispositions can be modified by differential reinforcement, and the second is that of whether changes in dispositions with regard to one dimension resulting from differential reinforcement restricted to that dimension will generalize to other dimensions in a manner consistent with their correlations with that dimension. The answer to the first question bears upon the applicability of the traditional principles of learning to the realm variously referred to as impression formation, person perception, or social perception. The answer to the second question, while sharing with the first this same general implication, goes further in having a direct bearing upon the validity and interpretation of the correlations between dimensions reported by Levy and Dugan.

The correlation between judgments for two traits obtained by the constant error method has been interpreted as indicating that they are not entirely independent of each other from the standpoint of perceptual functioning. That is, judgments for Trait A are believed to correlate to a greater extent with those for Trait B than they do with those for Trait C, because of those determinants governing each of these judgments, judgments for Traits A and B share a larger number in common than do judgments for Traits A and C. If this is so, then it should follow that modifications in judgmental tendencies in connection with one trait should be accompanied by modifications in judgmental tendencies in connection with other correlated traits. Further, it is reasonable to expect that the extent to which changes occur in judgments on these other traits should be directly related to the magnitude of their correlation with the initial trait for which judgmental tendencies have been modified.

[1] This investigation was supported in part by a Research Grant (M-2067) from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service.

[2] The author wishes to express his gratitude to Mary C. Skowronek who ran the subjects for this experiment.

It should thus become possible, granting the validity of the above reasoning, to order traits along a continuum on the basis of their correlations with a particular trait for which judgments are to be differentially reinforced, and to produce a gradient for the extent of changes in judgments for each of these traits which varies monotonically with the magnitudes of their correlations with the trait in question. Consequently, the following two hypotheses were proposed for testing in the present experiment:

Hypothesis I. The probability of making a particular judgment for a facial photograph on a given bipolar dimension is dependent upon the nature of the reinforcements contingent upon making such a judgment.

Hypothesis II. The relative magnitude of the probability of making a particular judgment, on a given bipolar dimension for which no previous reinforcement has been given, is directly related to the magnitude of the correlation between that dimension and the dimension for which judgments were differentially reinforced.

METHOD

Subjects. Subjects for this experiment were 30 men and 30 women, all students in an introductory psychology course which had as a course requirement participation in a certain number of experiments. They were randomly and evenly assigned to one of three groups with the only restriction being that there be an equal number of men and women in each group.

Stimuli. The facial photographs used in this experiment were of white, male applicants for graduate training in psychology at Indiana University who, for various reasons, did not enter the program. These pictures have been used in previous studies (Levy, 1960, 1961; Levy & Dugan, 1960) and therefore norms are available for each picture concerning the frequency with which a particular dichotomous judgment is made for it on a particular dimension. These norms have been found to be quite stable from one subject sample to another (Levy, 1961). Norms used in the selection of the present pictures were based on judgments made by 104 subjects for each of 75 pictures with regard to a particular dimension. The dimensions used for these judgments are those used in the present study; they are: safe-dangerous, good-bad, reliable-unreliable, and relaxed-tense. For the present study only pictures having values ranging between 40% and 60% for frequencies of judgments at one pole or the other were used. Twenty-two safe-dangerous pictures were used having a mean of 51.4% for "safe" judgments, five pictures each were used for each of the other dimensions having means of 50.2%, 50.6%, and 50.4% for judgments of "good," "reliable," and "relaxed," respectively.

Each picture was mounted on a 3 × 4 inch piece of poster board. Attached immediately below was a label containing the two terms designating the dimension on which that picture was to be judged. To control for possible position response sets, the left-right relationship for terms for approximately half of the pictures in each set was reversed.

Dimensions. The dimensions used in this study were chosen because each had been found by Levy and Dugan (1960) to be highly loaded on a different orthogonal factor. The correlations found by Levy and Dugan for each of the generalization dimensions with safe-dangerous were .50 for good-bad, .34 for reliable-unreliable, and .02 for relaxed-tense. While these correlations had been based on judgments by 100 subjects for only 15 pictures for each dimension, analogous correlations derived from the normative study involving 104 subjects and 75 pictures for each dimension were .80, .69, and .28, differing in absolute but not relative magnitude. Thus it would appear that some confidence may be placed in the reliability of a simple ordering of these dimensions with regard to the degree of their relationship to safe-dangerous.

Procedure. Safe-dangerous was selected as the dimension on which judgments would be differentially reinforced and the three groups of subjects were defined on the basis of the reinforcement combination used. The experimenter responded with "right" after all "safe" judgments and with "wrong" after all "dangerous" judgments for group SRDW, he responded in reverse fashion for group SWDR, and made no verbal response following judgments for the control group.

On Trials 1–20, which were the training trials, only pictures to be judged on the safe-dangerous dimension were presented. Trials 21–39 were test trials for generalization and on these a fixed random order was used in assigning the picture-judgment dimension for a given trial with the exception of Trials 24, 28, 32, and 36 on which safe-dangerous pictures were presented for the purpose of reducing any possible extinction effects of the generalization series.

Although the order of judgment dimensions over trials was the same for all subjects, the actual picture to be judged on a given trial was randomly selected for each subject from the pool of pictures assigned to the designated dimension. Thus, for example, on Trial 21, all subjects were required to make a judgment of

TABLE 1

ANALYSIS OF VARIANCE FOR CONDITIONED AND GENERALIZED CHANGES IN JUDGMENTS FOR FOUR DIMENSIONS

Source	*df*	*MS*	*F*
Between groups	2	111.713	55.004**
Subjects in same group	57	2.031	
Between dimensions	3	1.782	1.390
Dimensions × Groups	6	4.657	3.630*
Pooled subjects × Dimensions	171	1.282	

* $p < .01$.
** $p < 001$.

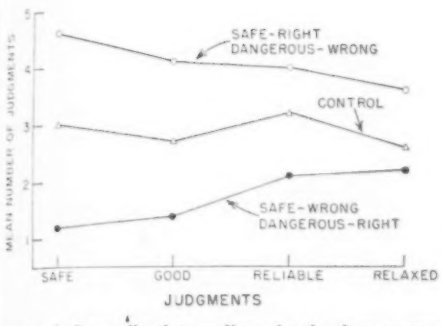

FIG. 1. Generalization gradients for the three groups of subjects for the judgments indicated on the abscissa.

"good" or "bad" although which of the five pictures assigned to this dimension might be presented on this trial varied randomly from subject to subject. No picture was presented more than once to a subject with the exception of the safe-dangerous ones of which there were only 22 available and 24 needed. Thus on two of the booster trials during the test series, the experimenter randomly selected and presented two pictures which had been previously presented to the subject.

The experiment was individually administered to each subject who was given the following instructions verbatim:

> This is one of a series of experiments pertaining to character evaluation. I am going to show you a number of pictures. Beneath each of these pictures are two descriptive words. Most of the descriptive words used are "safe-dangerous," but others will be shown to you intermittently. I want you to look at the pictures carefully; look at the words printed below the pictures, and then tell me which of the two words you think best describes this person.

Rate of presentation of the pictures was determined only by the time taken by the subject to report his judgment and the experimenter's recording of it. There was no apparent distinction from the subject's standpoint between the training and test trials.

RESULTS

Each subject received four scores: one for the number of "safe" judgments for Training Trials 16-20, which was taken as the measure of level of conditioning, and one each for the numbers of "good," "reliable," and "relaxed" judgments for the appropriate pictures, which were taken as the measures of generalization. Judgments for the safe-dangerous pictures used on booster trials were not analyzed.

Treating these four scores as repeated measures on the same individuals, an analysis of variance (Edwards, 1950) was performed that is summarized in Table 1. The F between groups, which is significant at the .001 level, supports the first hypothesis regarding the contingency between judgments and their reinforcements. Judgments made by subjects in this experiment were susceptible to conventional conditioning procedures.

The F for the interaction between dimensions and groups, which was significant at the .01 level, indicates that the trends of the means over the four dimensions for the three groups differ. As may be seen from the curves for the three groups plotted in Figure 1, judgments of pictures on the safe-dangerous dimension are affected by differential reinforcement in the expected way, and this effect does generalize to judgments on the other dimensions in a way that is monotonically related to the magnitude of their correlation with safe-dangerous. Thus the second hypothesis is also supported by these data.

DISCUSSION

The constant error method provides a means of studying the structure of social perception from the standpoint of perceptual dispositions. The present study has shown that such dispositions are susceptible to modification by differential reinforcement, and that such modifications ramify throughout the perceptual system in a manner consistent with the structure of this system as delineated by the constant error method. Thus at one and the same time, the results of the present study argue for the validity of the structural picture that emerges by the use of the constant error method, and also serve to qualify social perception for membership within the realm of phenomena to which current principles of learning have been found applicable.

But beyond this, the present findings are consistent with Wishner's (1960) demonstration that in the Asch-type (Asch, 1946) impression formation study changes in impressions as a result of changes in informational input follow lines predictable from empirical knowledge of the characteristics of the elements involved and their intercorrelations. This would suggest that given additional studies of the present type, as well as those of Asch and Wishner, the structure of the social perceptual system should become sufficiently clear so that it should ultimately be possible

to predict what effect any given bit of informational input into one part of the system has upon any other part.

The present study used what might be considered neutral stimulus objects with regard to the dimensions involved, and therefore leaves unanswered the question of the interaction between perceptual dispositions and values of the stimulus objects involved. Every judgment might be thought of as the resultant of this interaction and thus the question arises of the relative contributions of dispositions and objects to the final outcome. To answer this question, research incorporating an object sampling approach of the sort advocated by Brunswik (1947) is now being planned.

The predicted changes in perceptual dispositions observed in the present study were predicated upon the assumption of certain relationships existing between these dispositions. That these predictions were borne out supports the validity of this assumption. Now, another question of both theoretical and practical import is whether these very relationships can be altered. Conceivably, if they reflect the learning history of the individual with regard to situations requiring these judgments to be made, we should expect to find developmental changes in the structure of social perception, and we should be able to modify this structure by appropriate experimental treatment. Concretely, for example, the question is whether it is possible to reduce the correlation between judgments of safeness and goodness and increase that between safeness and relaxedness. Some approaches to psychotherapy assume that this can be done (e.g., Kelly, 1955), and the whole question of the possibility of modifying social stereotypes would seem to hinge upon this issue. Thus the question becomes one of the extent to which, and the means by which the relationships existing between perceptual dispositions may be modified. How, or whether, learning theory can make a contribution here remains to be determined.

A question that is undoubtedly begged by the present study, one that plagues much of the research in the area of perception, is whether the changes observed truly represent changes in perception or merely changes in verbal behavior. While the legitimacy of this question cannot be gainsaid, answering it in the context of the present study would not have been possible. For even if post-experimental interviews were conducted in order to determine whether subjects attributed veridicality to their judgments of pictures as either safe or dangerous, we would still not have gone beyond the level of verbal behavior. To properly decide between these two alternatives would require tapping into additional response systems of the organism and searching for correlations between data obtained from these systems (Garner, Hake, & Eriksen, 1956). Thus for the present, it must be recognized that an accounting for the present findings in terms of changes in verbal behavior remains a possible alternative.

SUMMARY

Subjects divided into two experimental and one control group were presented with sets of facial photographs for each of which they were to make a dichotomous judgment with regard to a specified bipolar dimension. Subjects in the experimental groups were conditioned to make either "safe" judgments or "dangerous" judgments on the first set of 20 pictures by means of differential reinforcement for each of these judgments, and then were tested for the extent to which the effects of this conditioning generalized to judgments for other pictures on good-bad, reliable-unreliable, and tense-relaxed dimensions. The nature of the generalization gradient had been predicted on the basis of correlations obtained in a previous study between judgments on each of these dimensions and those for the safe-dangerous dimension.

Differences between groups were statistically significant and the generalization gradient followed the expected course. The results were discussed from the standpoint of their implications for the nature of social perception and were interpreted as demonstrating the applicability of principles of learning to problems in this area.

REFERENCES

ASCH, S. E. Forming impressions of personality. *J. abnorm. soc. Psychol.*, 1946, **41**, 258–290.

BRUNSWIK, E. *Systematic and representative design of psychological experiments.* Berkeley: Univer. California Press, 1947.

EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.

FESHBACH, S., & SINGER, R. D. The effects of fear arousal and suppression of fear upon social perception. *J. abnorm. soc. Psychol.*, 1957, **55**, 283–288.

GARNER, W. R., HAKE, H. W., & ERIKSEN, C. W. Operationism and the concept of perception. *Psychol. Rev.*, 1956, **63**, 149–159.

HAYS, W. L. An approach to the study of trait implication and trait similarity. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior.* Stanford Univer. Press, 1958. Ch. 19.

KELLY, G. A. *The psychology of personal constructs.* New York: Norton, 1955.

LEVY, L. H. Context effects in social perception. *J. abnorm. soc. Psychol.*, 1960, **61**, 295–297.

LEVY, L. H. Adaptation, anchoring, and dissipation in social perception. *J. Pers.*, 1961, **29**, 94–104.

LEVY, L. H., & DUGAN, R. D. A constant error approach to the study of dimensions of social perception. *J. abnorm. soc. Psychol.*, 1960, **61**, 21–24.

MURRAY, H. A. The effect of fear upon estimates of maliciousness of other personalities. *J. soc. Psychol.*, 1933, **4**, 310–329.

NEWCOMB, T. An experiment designed to test the validity of a rating technique. *J. appl. Psychol.*, 1931, **22**, 279–289.

THORNDIKE, E. L. A constant error in psychological rating. *J. appl. Psychol.*, 1920, **4**, 25–29.

WISHNER, J. Reanalysis of "impressions of personality." *Psychol. Rev.*, 1960, **67**, 96–112.

(Received December 5, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 63, No. 3, 588-596

THE INFLUENCE OF AFFECTIVE FACTORS ON CONCEPTUALIZATION IN SCHIZOPHRENIA[1]

MELVIN H. FEFFER
Yeshiva University

A number of divergent positions exist with regard to the interpretation of pathological concreteness in certain types of schizophrenia. The viewpoint underlying the present study is that personality disorganization in schizophrenia is basic to pathological concreteness. This position can be articulated in terms of the following propositions:

Pathological concreteness of the schizophrenic is one manifestation of a generalized withdrawal from the affective disorganization engendered within interpersonal relationships (Arieti, 1955; Hunt & Cofer, 1944; Jenkins, 1950). Standard conceptual tasks have interpersonal implications not only by virtue of the social context within which these are typically administered, but, more importantly, in terms of the view that conceptual activity in itself requires a coordination of different social perspectives (Mead, 1934; Piaget, 1950). The schizophrenic's focusing on peripheral, irrelevant aspects of a situation (Cameron, 1939; Von Domarus, 1944; Wegrocki, 1940) is interpreted as one technique of withdrawing from the interpersonal and hence affect arousing implications of the conceptual task. More formally, one can hypothesize that the schizophrenic avoids the central, affect arousing implications of the conceptual task by focusing upon peripheral, irrelevant material, thereby evidencing inadequate conceptual functioning in terms of standard test criteria.

A direct test of this hypothesis would involve varying affective and neutral components of a conceptual task along a "central-peripheral" dimension. Standard tests of conceptual activity, however, lend themselves less readily to such manipulation than perceptual figure-ground tasks, which can be so designed that contents in both the central (figure) and peripheral (ground) dimensions are either affective or neutral. The use of a figure-ground task in a test of the hypothesis, however, requires an assumption that pathological concreteness and inadequate figure-ground differentiation represent similar activities. In this regard, Goldstein (1939) has pointed out that the lack of subordination of ground to figure which occurs in inadequate figure-ground differentiation also characterizes pathological concreteness. That is, the pathologically concrete attitude is one in which each object in the environment is considered as unique and self-contained and, hence, as important as every other object, rather than as subordinate exemplars of more general categories.

Assuming that inadequate figure-ground differentiation and pathological concreteness are similar in terms of lack of subordination, one may advance the following experimental hypothesis: Pathologically concrete schizophrenics will evidence an avoidance of affective stimuli in figure-ground differentiation when figure and ground are varied in terms of affective and neutral valence. More specifically, the writer predicts that (*a*) under affective figure conditions, pathologically concrete schizophrenics focus on ground aspects of the field; (*b*) under affective ground conditions, pathologically concrete schizophrenics focus on figure aspects of the field; and (*c*) these patterns of behavior will be more characteristic of pathologically concrete schizophrenics than of two control groups—adequately conceptualizing schizophrenics and normal individuals.

METHOD

Subjects and General Procedure

Forty-one male subjects who had normal or corrected vision were employed. These consisted of the following three groups[2]:

[1] This article is based upon a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology at the University of Chicago.

The author wishes to thank David Shakow for his guidance during this study and Donald Broverman who made many helpful suggestions with regard to the present paper.

This investigation has been supported in part by the Dementia Praecox Research Project, Worcester State Hospital, and a research grant (M-896) from the National Institute of Mental Health, United States Public Health Service.

[2] The patient population was drawn from Elgin State Hospital. The diagnoses were made by a staff

1. Concrete group: 9 schizophrenic subjects who showed pathological concreteness on criterion tasks of conceptual thinking.

2. Conceptualizing group: 16 schizophrenic subjects who showed adequate conceptualization on criterion tasks of conceptual thinking.

3. Normal group: 16 subjects who, as far as could be ascertained, had no history of hospitalization for mental illness. Eleven subjects in this group were individuals working in the community and living at the YMCA. They were paid for taking part in the study. The rest of the group consisted of graduate students or professional workers in psychology.

All subjects were seen in three sessions. In the first or screening session, the tests administered were the Stanford-Binet Vocabulary (1937) and two tests of conceptual ability derived from Wegrocki (1940). The purpose of this screening period was to select those subjects in the schizophrenic group who were extreme cases on the continuum of concrete-abstract thought, and to rule out those subjects who gave the impression that cooperation would be minimal on the more effort involving tasks. Following a random time interval of 1-7 days, each subject was given a figure-ground tachistoscopic task. A third session followed a time interval of 3 days in which the Luria (1932) (word association-motor response) task, and a Wittenborn (1943) test of concentration were given. This latter test was used as a check on the effectiveness of the screening period in yielding cooperative subjects.

The figure-ground situation was so constructed that figure and ground could be varied in terms of affective and neutral valence. This was accomplished by the use of a tachistoscopic task in which meaningful words were placed in a central area on the stimulus field, while other meaningful words were placed on the periphery. All words were classified as either affective or neutral for each subject on the basis of the word association-motor technique developed by Luria (1932). Subjects were required to identify the central word in the figure-ground situation. A measure was devised which represented the presence of background content in the response to the figure, i.e., "ground intrusion." Relationships between the degree of ground intrusion and the locus of the affective word—whether in figure or in ground—were then analyzed in order to determine whether avoidance of affective words had taken place.

Criterion Tasks of Conceptual Thinking

The tasks used as the basis for classifying each schizophrenic subject as pathologically concrete or as an adequate conceptualizer were the Proverb Interpretation test and the Essential Differences test (Wegrocki, 1940). The Proverb Interpretation test consists of two groups of proverbs and one group of interpretations. Each proverb in the first group corresponds in meaning to a proverb in the second group, although worded differently, e.g., "When in Rome, do as the Romans do" (Group 1) and "When in a forest, one must howl with the wolves" (Group 2). The subject is first required to match the proverbs in the first two groups, and then he is asked to match the proverbs in the first group with the interpretations in the third group, e.g., "If one wants to live on good terms with his neighbors one must act as they do." Following Wegrocki, the subject's score is based on the number of errors made. Each incorrect matching of proverbs is given one point, while each incorrect matching of proverb and interpretation is given two points. On the basis of Wegrocki's study it was decided that those schizophrenics scoring 9 or under would be classified as low error schizophrenics, while those scoring over 18 would be classified as high error schizophrenics.

The Essential Differences test consists of groups of words, four words to a group. Three of the four words are similar in some respect while the fourth is not, e.g., "chimney, *fire*, smokestack, stovepipe." The subject is required to underline the word that differs from the remaining three. The following instructions are read to each subject:

> I'd like you to underline that word in each line which you think is most different from the remaining three; but the words left must be in some way related and similar. For example, of these four words, chimney, fire, smokestack, stovepipe, which would you underline?

The subject's score is based on the number of errors made, each incorrect underlining or omission being given one point. Based on Wegrocki's findings, a critical score of 10 was chosen. Those schizophrenics scoring over 10 were classified as high error schizophrenics while those scoring 10 or under were classified as low error schizophrenics. For the present study, only those schizophrenics who were high error scorers on both tests were placed in the Concrete group. Similarly, only those schizophrenics who were low error scorers on both tests were placed in the Conceptualizing group. None of the Normal group exceeded the critical scores on either task. There is no significant difference in scoring between the Conceptualizing group and the Normal group.

TABLE 1

RELIABILITY COEFFICIENTS OF PROVERB INTERPRETATION, ESSENTIAL DIFFERENCES, AND WITTENBORN CONCENTRATION TESTS

Item	Test				
	Proverb Interpretation	Essential Differences	Wittenborn Concentration		
			N-C	S-A	S-C
Group[a]	N-C	N-C			
r	.82	.56	.88	.82	.50
SE	1.96	1.66	2.17	3.50	7.82

[a] N-C: normal control group; S-A: adequately conceptualizing schizophrenic group; S-C: pathologically concrete schizophrenic group.

of psychiatrists and psychologists, or by the psychiatrist in charge of the ward to which the patient was assigned.

Criteria of Affective and Neutral Words: Luria Task

An estimate of the affective arousal of the words used in the tachistoscopic task was based on the Luria task (word association-motor response). The method and rationale are described by Luria (1932), who pioneered work in this area. Validation studies have been presented by Hull and Lugoff (1921), Huston, Shakow, and Erickson (1934), and Sharp (1938); other studies have shown that the test can be used with psychotic populations (Ebaugh, 1936; Shuey, 1937).

The subject is seated in a comfortable chair while fingers of both hands rest on receiving tambours (rubber cups). He is asked to press down on one rubber cup with his dominant hand as he verbalizes his first association to the examiner's stimulus word, leaving his other hand resting lightly on the other rubber cup. Movements in either hand and reaction time were recorded on separate lines of a Phipps-Bird kymograph.

The motor response of each subject was analyzed by the investigator without knowledge of either the stimulus word that elicited this response or the content of the subject's association. A disturbance of either reaction time or dominant hand activity was sufficient to put that stimulus word in the affective category for that individual. If neither was considered disturbed, the word was categorized as neutral. A more complete description of the technique of assessing the motor responses is given in Luria (1932), and Feffer (1954).

Reliability of Tests

The reliability coefficient of the tests used may be seen in Table 1. These were determined by the split-half method, using Rulon's method (Guilford, 1942, p. 276) for the Essential Differences and the Wittenborn Concentration tasks. The reliability coefficient for the Proverb Interpretation task was determined by the alternate forms method, using rank-order correlation. Since the subject's score was determined by both forms, the Spearman-Brown formula was used.

Reliability coefficients for the Essential Differences and the Proverb Interpretation tasks are based on the scores of the normal controls only. This is due to the fact that each schizophrenic group was, by design, selected for extreme homogeneity.

Interjudge Reliability

Seven representative Luria records were selected. Random sections of 10 consecutive responses which the experimenter and a second judge classified as affective or neutral were selected from each record. The degree of agreement between judges for each block of 10 consecutive responses was determined by the use of a contingency table. The null hypothesis was tested each time. Exact probabilities derived from each contingency were then combined in order to obtain an over-all probability measure. On the basis of such a test of significance, the hypothesis that the degree of agreement could have arisen by chance can be rejected at the .001 level. The median agreement was 90%.

Figure-Ground Tachistoscopic Task

The instructions to the subject were read orally:

A number of words are going to be shown to you. They are going to be shown very quickly so you will have a great deal of difficulty in seeing them. I want you to tell me in each case what the single diagonal word in large type is. It is best if you look at the place where the "X" is, because that is where the word will appear. Do not pay any attention to horizontal words in small type. Since you will have difficulty in seeing the diagonal word, be sure to guess if you are not sure.

Sufficient time was taken on the first few practice trials to insure that the subject understood the nature of the task. There were 118 trials in this task; the first 18 were practice trials and were not counted in the score. The order of presentation was the same for each subject. Each trial was presented at 1/5 second, with a verbal warning 2 seconds prior to the presentation of the figure-ground stimulus.[3] A Gerbrands modified Dodge tachistoscope was used for the presentation of the figure-ground stimuli. An example of the stimulus field is given below:

sugar A sugar
sugar P sugar
sugar P sugar
sugar L sugar
sugar E sugar

Figure and ground are defined by the "set" given in the instructions (Werner, 1949). Thus, it is assumed that when the examiner stresses that the subject report the central diagonal word, this part of the stimulus field becomes the figure or the goal toward which activity is directed. It is on this basis that the diagonal word is termed figure and the surrounding words ground.

Figure and ground words were so combined that there was no duplication of *consonants* in the figure and ground for any given trial. For example, if the ground word contained the consonants "s," "g," and "r," the experimenter selected a word for the figure which contained none of these letters. Such a procedure, however, could not be followed for vowels.

The following are possible responses to the example given above (ground = "sugar," figure = "apple").

1. "anger"	*G* (ground intrusion: consonants "r" and "g" in the ground as well as in the incorrect response to figure)	Incorrect Responses
2. "avid"	*X* (incorrect response to figure in which no consonants of the ground appear)	
3. "missed that one"	*NR* (no response)	
4. "apple"	*C* (Correct response)	

Two measures were derived from these responses. The first is *%Inc* (number of incorrect responses divided by the total number of figure-ground presenta-

[3] The preparatory time of 2 seconds was based on the study by Huston, Shakow, and Riggs (1937) in which it was found that schizophrenic subjects tended to take longer periods of time to reach their peak of readiness to respond than did normal subjects. A preparatory time of 2 seconds did not appear to penalize one group more than the other.

TABLE 2
FIGURE-GROUND AVOIDANCE COMPARISONS

Figure	Ground	Figure	Ground	Avoidance measures
affective	neutral	neutral	neutral	Avoidance of figure: figure varies, ground constant
neutral	affective	neutral	neutral	Avoidance of ground: figure constant, ground varies
affective	neutral	neutral	affective	Extreme avoidance comparison

tions). This score is considered an index of general adequacy of figure-ground differentiation.

The second measure is based on the ground intrusion response which represents the degree to which ground content is present in the response to figure. It should be emphasized at this point that all subjects were given the same figure-ground stimuli. Depending on the subject's Luria responses, however, a given figure-ground situation might be affective-figure:neutral-ground for one subject, and neutral-figure:affective-ground for another subject. For this reason, scores are reported as percentages, i.e., *%G*, the number of ground intrusions in a given classification divided by the total number of presentations of that classification.

Criteria of Avoidance

Since the words used in figure or ground were classified as either affective or neutral for each subject, the following classification of figure-ground stimuli is possible: neutral-figure:neutral-ground, neutral-figure:affective-ground, affective-figure:neutral-ground, and affective-figure:affective-ground. The first three of these classifications, as seen in Table 2, were used in assessing avoidance.

Avoidance was defined as either larger *%G* under affective-figure:neutral-ground conditions than under neutral-figure:neutral-ground conditions, smaller *%G* under neutral-figure:affective-ground conditions than under neutral-figure:neutral-ground conditions, or larger *%G* under affective-figure:neutral-ground conditions than under neutral-figure:affective-ground conditions.

RESULTS

Within-group comparisons of *%G* under the various figure-ground situations are presented in Table 3. It may be seen that for the Concrete group, there is a larger *%G* under affective-figure:neutral-ground conditions than under neutral-figure:neutral-ground conditions, a smaller *%G* under neutral-figure:affective-ground conditions than under neutral-figure:neutral-ground conditions, and a larger *%G* under affective-figure:neutral-ground conditions than under neutral-figure:affective-ground conditions. These data confirm the predictions that pathologically concrete schizophrenic subjects focus on ground aspects of the situation under affective figure conditions and focus on figure aspects of the situation under affective ground conditions. Accordingly, it may be inferred that the findings support the hypothesis than an avoidance of affective stimuli is evidenced in the figure-ground differentiation of pathologically concrete schizophrenics when figure and ground are varied in terms of affective and neutral valence.

The prediction that such avoidant behavior is more characteristic of the performance of the Concrete group than either the Conceptualizing or the Normal groups is also supported by the present findings. As may be seen in Table 3, there are no significant differences between the various figure-ground conditions in the Normal group. On the other hand, one may see a pattern of relationships in the performance of the Conceptualizing group *opposite* to that present in the Concrete group. Thus, for this group, there is a larger *%G* under neutral-figure:affective-ground conditions than under neutral-figure:neutral-ground conditions, a smaller *%G* under affective-figure:neutral-ground conditions than under neutral-figure:neutral-ground conditions, and a larger *%G* under neutral-figure:affective-ground conditions than under affective-figure:neutral ground conditions.

In summary, the Concrete group meets the criteria of avoidance in the three comparisons. The Conceptualizing group not only

TABLE 3
WITHIN-GROUP COMPARISON OF %G

Group	Figure-ground comparisons[a]	p[b]
Concrete	NA<NN	.05
	AN>NN	<.10>.05
	NA<AN	.02
Conceptualizing	NA>NN	<.10>.05
	AN<NN	.10
	NA>AN	<.01
Normal	NA<NN	>.25
	AN<NN	>.25
	NA<AN	>.25

[a] AA: affective-figure:affective-ground; AN: affective-figure:neutral-ground; NA: neutral-figure:affective-ground; NN: neutral figure:neutral-ground.

[b] The Wilcoxon (1948) *T* test was employed for the comparisons in this table. The significances reported are two-tailed.

fails to meet the criteria of avoidance in any of the comparisons, but these subjects evidence a pattern of behavior opposite to that predicted on the basis of avoidance and, consequently, opposite to that of the Concrete group. The Normal group neither meets the criteria of avoidance in any of the comparisons nor shows behavior patterns opposite to those predicted on the basis of avoidance.

In order to explore further the degree to which the three groups differ in avoidance of affective stimuli, the three comparisons involving *%G* were analyzed for each subject. Each subject was given a plus for each of the comparisons in which the difference in *%G* conformed to the criterion of avoidance of affective stimuli and was given a minus for each of the comparisons in which the difference in *%G* did not conform to such a criterion. For example, if a given subject had a larger *%G* in the affective-figure:neutral-ground situation than in the neutral-figure:neutral-ground situation, he was assigned a plus; if he had a smaller *%G* (or equal) under affective-figure:neutral-ground conditions than under neutral-figure:neutral-ground conditions, he was assigned a minus. A test of significance of the incidence of plus and minus scores for each of the three comparisons was made among the three groups. If one group had a significantly greater incidence of plus than another group for any given comparison, this was taken as indicating that, for this particular comparison, the first group showed greater avoidance of affective stimuli. Table 4 shows the pertinent relationships. In the three comparisons, the Concrete group showed significantly greater avoidance of affective stimuli than the Conceptualizing group. It will also be noted that in one comparison, the Normal group showed a greater avoidance than the Conceptualizing group. A composite avoidance score for each subject was also derived by computing a plus percentage (plus/plus + minus). It may be seen that the Concrete group had the highest and the Conceptualizing group the lowest composite avoidance scores. The Normal group fell between these extremes.

Controls

In order to determine whether the obtained relationships could be attributed to the operation of variables other than those hypothesized, the following characteristics were investigated: age, educational level, vocabulary level, level of cooperation as indicated by Wittenborn scores, length of hospitalization, and word frequency. As shown in Table 5, there are no significant differences among the three groups in age, educational level, Wittenborn scores, and no significant difference between the schizophrenic groups in length of hospitalization. The Concrete group is significantly lower in vocabulary level than either the Conceptualizing group, or the Normal group. However, no significant difference obtains between the Conceptualizing group and the Normal group in vocabulary level.

In controlling for the higher vocabulary performance of the Conceptualizing group, the investigator dropped from this group the eight subjects with the highest vocabulary scores. He then compared this restricted group to the Concrete group from which the subject with the lowest vocabulary score was omitted. The difference in vocabulary level between these two restricted groups was not significant.

TABLE 4

BETWEEN-GROUP COMPARISON ON AVOIDANCE OF AFFECTIVE STIMULI GROUP COMPARISONS[a]

Avoidance criteria[b]	SC/SA	NC/SA	SC/NC
NN > NA	SC > SA (.01)	—	—
AN > NN	SC > SA (<.01)	—	—
AN > NA	SC > SA (<.01)	NC > SA (<.01)	—
Composite[c]	SC > SA (<.01)	NC > SA (<.01)	SC > NC (.05)

Note.—The exact method for computing probabilities for 2 × 2 contingency table was employed for the comparisons in this table. The significances reported are two-tailed.

[a] SC: pathologically concrete schizophrenic group; SA: adequately conceptualizing schizophrenic group; NC: normal control group.

[b] AA: affective-figure:affective-ground; AN: affective-figure:neutral-ground; NA: neutral-figure:affective-ground; NN: neutral-figure:neutral-ground.

[c] The Mann-Whitney (1947) *U* test was employed in comparing group on the composite score. The significances reported are two-tailed.

TABLE 5

CHARACTERISTICS OF THE PATHOLOGICALLY CONCRETE SCHIZOPHRENIC, THE ADEQUATELY CONCEPTUALIZING SCHIZOPHRENIC, AND THE NORMAL CONTROL GROUPS

Group	Variable									
	Age (Years)		Educational level (Years)		Stanford-Binet Vocabulary (Words completed)		Wittenborn Concentration (Number of errors)		Length of hospitalization (Number of months)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Pathologically concrete schizophrenic (S-C) ($N = 9$)	32.6	3.13	10.9	2.28	22.9	5.36	11.9	8.62	36.6	10.7
Adequately conceptualizing schizophrenic (S-A) ($N = 16$)	29.6	5.66	11.9	2.08	29.2	4.28	12.2	8.37	22.8	39.3
Normal Control (N-C) ($N = 16$)	31.1	7.09	13.9	4.05	28.1	5.57	12.2	10.6	—	—
SC/SA	1.40		1.00		3.09[a]		.08		.46	
SC/NC	.58		1.97		2.18[b]		.07		—	
SA/NC	.64		1.71		.61		.00		—	

[a] At or beyond the 1% level of significance.
[b] At the 5% level of significance.

However, differences in avoidance scores between the two restricted groups continue to be significant in spite of the smaller number of subjects.

The experimenter dealt with the problem of word frequency by using words that occurred within the first 6,000 most frequently used words (Thorndike & Lorge, 1944), rather than socially tabu words having a low frequency of occurrence. However, in spite of employing a narrow range of word frequency, it is possible that the relationships obtained between affectivity in the figure words and the phenomenon of ground intrusion may still be accounted for by differences in the frequency of occurrence of those words judged affective and those words judged neutral. The fact that the judgment of affective arousal is based on Luria criteria for each individual does not meet a "frequency criticism" because the question may be raised of whether less familiar words might be more difficult for the individual to associate to, result in disorganization on the Luria, and hence be judged affective. In order to investigate this possibility, a test of significance by means of the χ^2 test of ranks (Wilcoxon, 1948) was made between frequency of occurrence of the words used as figure and the judgment of affective arousal by Luria criteria. No significant relationship, however, was found between word frequency and the judgment of affectivity by Luria criteria. The problem of word frequency does not exist for the relationships between background variation in affectivity and ground intrusion since all words in the background were of the same frequency of occurrence. Thus, the variable of familiarity as defined by the Thorndike-Lorge count can be ruled out as accounting for present findings. It is obvious, however, that individual subject differences in familarity with the test words are not thereby discounted.

DISCUSSION

The results of the present study provide support for the experimental hypothesis that an avoidance of affective stimuli is evidenced in the figure-ground differentiation of pathologically concrete schizophrenics. It is less clear, however, as to whether the findings also provide support for the general hypothesis that pathologically concrete thinking in schizophrenia represents an avoidance reaction to affective stimuli. It will be recalled that the use of the figure-ground perceptual task is based on the assumption that inadequate figure-ground differentiation and pathologically concrete thinking can be equated. A check on this assumption entails a comparison of the three groups with regard to the number of incorrect responses in the figure-ground task (*%Inc*). When the three groups are compared in over-all performance by means of the Mann-Whitney *U* test, no significant differences are present in *%Inc*.

In addition, there are no significant differences between groups in *%Inc* under any of the specific figure-ground conditions. In terms of this measure, therefore, there is no confirmation that pathologically concrete schizophrenics show less adequate figure-ground differentiation than either adequately conceptualizing schizophrenics or normal individuals. The empirical basis for equating inadequate figure-ground differentiation and pathologically concrete thinking is thereby lacking; accordingly, affect avoidance as inferred from the schizophrenic's figure-ground differentiation cannot be attributed to his pathologically concrete thinking. The results may therefore be interpreted as being consistent with, rather than as directly supportive of, the general hypothesis that pathologically concrete thinking in schizophrenia represents an avoidant reaction to affective stimuli.

It is possible that a methodological weakness such as insufficient sensitivity of a measuring instrument can account for the lack of difference between groups in adequacy of figure-ground differentiation. This consideration may be particularly relevant in view of the fact that "nontabu" words were employed, i.e., those occurring within the 6,000 most frequently used words (Thorndike & Lorge, 1944). Of more theoretical import, however, is a reconsideration of the assumption that inadequate figure-ground differentiation and pathological concreteness represent similar activities. It will be recalled that the schizophrenic subjects were selected on the basis of cooperation as well as extreme performance on tasks of conceptual ability. Shakow (1946) in presenting experiments investigating schizophrenic behavior at different levels of complexity has concluded that once degree of cooperation is controlled, disturbances of psychological function are not seen at lower levels of complexity such as the pattelar tendon reflex, but are only evidenced at more complex levels of behavior such as conceptualization. The present lack of group differences in adequacy of figure-ground differentiation suggests the possibility that the figure-ground tasks differ from the conceptual tasks in terms of degree of complexity, and that lack of subordination as such is an insufficient basis for equating inadequate performance on these tasks. This reappraisal does not negate the significance of the finding that pathologically concrete schizophrenics evidence avoidance of affective stimuli in the figure-ground task. It does, however, raise the possibility that avoidance of affect arousal may take quite different forms as a function of the complexity-level of behavior analyzed. If conceptual activity is viewed as involving the simultaneous coordination of different social perspectives (Mead, 1934; Piaget, 1950), then the avoidance of interpersonal anxiety may necessitate a withdrawal from the very activity of conceptualization itself because of its immanent interpersonal quality. Arieti (1955) has suggested a similar notion in his evaluation of the schizophrenic's perceptualization of thought as a "retreat from reason." If this were the case, then conceptual deficit would be present whether the task and its context were designated as affective, neutral, social, or impersonal. Perceptual differentiation, on the other hand, has been interpreted as involving a qualitatively different type of behavior, viz., a series of sequential focusing activities rather than simultaneous coordination (Feffer, 1959; Feffer & Gourevitch, 1960; Piaget, 1950). This type of activity could well be differentially influenced by the locus of affective and neutral stimuli. In view of these considerations, the equating of inadequate figure-ground differentiation and pathologically concrete thought may have also been incorrect in terms of attempting to fit conceptual thought within an affect avoidance paradigm operationally appropriate to behavior at a less complex level of organization.

The inference of affect avoidance in the Concrete group should also be considered in relation to the performance of the Conceptualizing group. The difference between the two groups in their reaction to affective stimuli was not, as anticipated, one of degree. The adequately conceptualizing schizophrenics, rather than merely showing less avoidance, appeared "drawn" to affective stimuli and, in fact, were significantly more extreme in this regard than the normal controls. Although unexpected, the diametrically opposite patterns of affective orientation evidenced by the two schizophrenic groups

may not be an isolated finding. The pattern of differential sensitivity to anxiety provoking stimuli termed vigilance and defense (Chodorkoff, 1954; Lazarus, Eriksen, & Fonda, 1951), phenomenally, at least, is similar to the present opposite affective orientations. Also pertinent to the present results may be studies in a rather different area of investigation: physiologic (including autonomic nervous system) responsiveness in schizophrenia (King, 1958; Malmo & Shagass, 1949; Meadow, Greenblatt, Funkenstein, & Soloman, 1953; Williams, 1953). From these studies, it seems reasonable to infer that conceptually adequate "reactive" schizophrenics are physiologically more responsive than both conceptually impaired "process" schizophrenics and normals. The pattern of physiologic responsiveness in relation to level of conceptual ability resembles in certain respects the directional sensitivity to affective material found in the present study, viz., the adequately conceptualizing schizophrenic's heightened positive sensitivity, the conceptually impaired schizophrenic's heightened negative sensitivity, and the normal individual's position between these extremes. Further studies could well be directed toward the nature of the relationship between adequacy of conceptual ability, level of physiologic responsiveness, and affect orientation, not only in schizophrenia but also, as the vigilance and defense studies suggest, in normal and neurotic groups as well. An important possibility is that lowered physiologic responsiveness in the conceptually impaired process schizophrenic represents not a simple response deficiency, but an active, ongoing, avoidant reaction as part of his generalized withdrawal from interpersonal relationships.

SUMMARY

The present study is directed toward a test of the hypothesis that pathologically concrete thinking in schizophrenia represents an avoidance reaction to affective stimuli. To this end, a figure-ground situation was constructed in which figure and ground were varied with regard to affective valence. A ground intrusion measure, representing the presence of background content in the response to figure was devised. The figure-ground differentiation of schizophrenic and normal individuals who were independently evaluated as to conceptual level, was analyzed with regard to relationships between ground intrusions and the locus of affective stimuli (whether in figure or in ground). The predictions were that (*a*) under affective figure conditions pathologically concrete schizophrenics would focus on ground aspects of the situation; (*b*) under affective ground conditions pathologically concrete schizophrenics would focus on the figure; and (*c*) these patterns of behavior would be more characteristic of pathologically concrete schizophrenics than of either adequately conceptualizing schizophrenics or normal controls. The results are interpreted as supporting these predictions and as being consistent with the main hypothesis.

REFERENCES

ARIETI, S. *Interpretation of schizophrenia.* New York: Brunner, 1955.

CAMERON, N. Deterioration and regression in schizophrenic thinking. *J. abnorm. soc. Psychol.*, 1939, **34,** 265–270.

CHODORKOFF, B. Self-perception, perceptual defense, and adjustment. *J. abnorm. soc. Psychol.*, 1954, **49,** 508–512.

EBAUGH, F. Association-motor investigation in clinical psychiatry. *J. ment. Sci.*, 1936, **82,** 731–743.

FEFFER, M. The influence of affective factors on conceptualization in schizophrenia. Unpublished doctoral dissertation, University of Chicago, 1954.

FEFFER, M. The cognitive implications of role-taking behavior. *J. Pers.*, 1959, **27,** 152–168.

FEFFER, M., & GOUREVITCH, VIVIAN. Cognitive aspects of role-taking behavior in children. *J. Pers.*, 1960, **28,** 383–396.

GOLDSTEIN, K. *The organism.* American Book, 1939.

GUILFORD, J. P. *Fundamental statistics in psychology and education.* New York: McGraw-Hill, 1942.

HULL, C. L., & LUGOFF, L. S. Complex signs in diagnostic free association. *J. exp. Psychol.*, 1921, **4,** 11–136.

HUNT, J. MCV., & COFER, C. N. Psychological deficit. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald, 1944. Pp. 971–1033.

HUSTON, P. E., SHAKOW, D., & ERICKSON, M. H. A study of hypnotically induced complexes by means of the Luria technique. *J. gen. Psychol.*, 1934, **11,** 65–97.

HUSTON, P. E., SHAKOW, D., & RIGGS, L. A. Studies of motor function in schizophrenia: II. Reaction time. *J. gen. Psychol.*, 1937, **16,** 39–82.

JENKINS, R. L. Nature of the schizophrenic process. *Arch. Neurol. Psychiat.*, 1950, **64,** 243–262.

KING, G. F. Differential autonomic responsiveness in

the process-reactive classification in schizophrenia. *J. abnorm. soc. Psychol.*, 1958, **56,** 160–164.

LAZARUS, R. S., ERIKSEN, C. W., & FONDA, C. P. Personality dynamics and auditory perceptual recognition. *J. Pers.*, 1951, **19,** 471–483.

LURIA, A. P. *Nature of human conflicts.* New York: Liverright, 1932.

MALMO, R. B., & SHAGASS, C. Physiologic studies of reaction to stress in anxiety and early schizophrenia. *Psychosom. Med.*, 1949, **11,** 9–24.

MANN, H. B., & WHITNEY, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **18,** 50–60.

MEAD, G. *Mind, self, and society.* Chicago: Univer. Chicago Press, 1934.

MEADOW, A., GREENBLATT, M., FUNKENSTEIN, D. H., & SOLOMAN, H. C. Relationship between capacity for abstraction in schizophrenia and physiologic response to autonomic drugs. *J. nerv. ment. Dis.*, 1953, **118,** 332–338.

PIAGET, J. *The psychology of intelligence.* New York: Harcourt, Brace, 1950.

SHAKOW, D. The nature of deterioration in schizophrenic conditions. *Nerv. ment. dis. Monogr.*, 1946, No. 70.

SHARP, D. L. Group and individual profiles in the association-motor test. *U. Iowa Stud. Child Welf.*, 1938, **15,** 97–196.

SHUEY, H. An investigation of the Luria with normal and psychotic subjects. *J. abnorm. soc. Psychol.*, 1937, **32,** 303–343.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Bureau of Publications, 1944.

VON DOMARUS, E. The specific laws of logic in schizophrenia. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia.* Berkeley: Univer. California Press, 1944. Pp. 104–113.

WEGROCKI, H. Generalizing ability in schizophrenia: An inquiry into the disorders of problem thinking in schizophrenia. *Arch. Psychol., N.Y.*, 1940, No. 254.

WERNER, H. Thought disturbance with reference to figure-background impairment in brain injured children. *Conf. Neurol.*, 1949, **9,** 255–263.

WILCOXON, F. *Some rapid approximate statistical procedures.* American Cynamid, 1948.

WILLIAMS, M. Psychophysiological responsiveness to psychological stress in early chronic schizophrenic reactions. *Psychosom. Med.*, 1953, **15,** 456–462.

WITTENBORN, J. R. Factorial equations for tests of attention. *Psychometrika*, 1943, **8,** 19–35.

(Received December 7, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 63, No. 3, 597-602

BASIC DIMENSIONS OF CHANGE IN THE SYMPTOMATOLOGY OF CHRONIC SCHIZOPHRENICS

JOHN E. OVERALL,[1] DONALD R. GORHAM
Veterans Administration Hospital, Perry Point, Maryland

AND JOHN R. SHAWVER
Veterans Administration Hospital, Waco, Texas

THE present analysis was undertaken in an effort to gain insight into the nature of the "schizophrenic process" by examining the functional covariation of symptoms as the psychiatric condition of the patient changes. Symptoms that depend upon a unitary psychiatric process should change together as the condition of the patient improves or worsens. If schizophrenia is such a unitary psychiatric process, one should be able to define it in terms of a set of symptoms that change together as the schizophrenic process develops. Symptoms that are related to the schizophrenic process should change independently of symptoms that depend primarily upon other psychopathologic processes.

Other investigators have attempted to characterize differences between psychiatric patients by examining differences in symptom patterns at one point in time. Eysenck (1956) has been concerned with basic dimensions of difference between different diagnostic groups of patients. Using the multiple discriminant function approach, he concluded that "psychoticism" represents one type of unidimensional deviation from normal and that "neuroticism" represents another distinctly different type of deviation. The symptoms that are useful in discriminating neurotics from normals thus may not be the same as those which are useful in discriminating psychotics from normals.

Lorr (1953) used methods of factor analysis to identify basic dimensions of difference among individual mental patients. He found differences in symptomatology to be multidimensional, with 11 somewhat independent dimensions being required for a comprehensive description of differences among individual hospitalized patients. It should be noted that the findings of Lorr and Eysenck need not be considered in conflict because different types of questions were being asked. Eysenck's approach was concerned with differences between diagnostic groups of individuals, whereas Lorr was concerned with differences among individuals, many of whom belonged to the same diagnostic groups.

In contrast to problems dealt with by other investigators, as illustrated by the studies mentioned, a distinctly different issue is considered in this article. What are the basic dimensions of individual difference in *change* in psychiatric symptoms for patients having a somewhat homogeneous psychiatric classification?

The methods of factor analysis were applied to the matrix of intercorrelations derived from change scores on 42 items[2] of the psychiatric interview portion of the Multidimensional Scale for Rating Psychiatric Patients (MSRPP) developed by Lorr (1953).[3] The data came from a study by Shawver, Gorham, Leskin, Good, and Kabnick (1959) in which three matched groups, each composed of 40 chronic schizophrenic patients who had been maintained on chlorpromazine for a period of 6 months or longer, were (*a*) continued on chlorpromazine, (*b*) shifted to maintenance on reserpine, or (*c*) shifted to placebo.

In the previously reported analysis of these data, interest was centered on the evaluation of differences between the three maintenance treatment regimens with the "total morbidity"

[1] Now in Department of Psychology, Kansas State University, Manhatten, Kansas.

[2] The psychiatric interview portion of the MSRPP contains 40 numbered items plus two additional subitems, 22a and 27a.

[3] In addition to the 42 items from the Lorr scale, five global ratings accomplished by ward personnel (Shawver, Gorham, Leskin, Good, & Kabnick, 1959) were included in the correlation matrix and factor analysis. Since these measures represent a completely different type of data, and since they were not found to correlate significantly with the rotated factors, they are not considered further in this report.

score from the MSRPP used as the major dependent variable. The treatment changes resulted in significant deterioration in the psychiatric condition of patients from whom chlorpromazine was withdrawn. Thus the data were thought to be especially suitable for the present analysis, which was undertaken to provide insight into the nature of *changes* in psychiatric pathology. Since the interest was in understanding basic dimensions of individual difference in psychiatric change, differences between drug groups were partialed out in order to remove this extraneous source of variance before the present analysis was undertaken.

TABLE 1

Loadings of 42 Items from Psychiatric Interview Portion of MSRPP on Eight Centroid Factors

	Ic	IIc	IIIc	IVc	Vc	VIc	VIIc	VIIIc
1.	.55	.33	.30	.06	.06	.13	−.22	−.16
2.	.16	.10	.41	.07	−.02	−.09	−.08	−.13
3.	.47	.11	.25	.17	.12	−.06	−.13	−.13
4.	.39	.25	.22	.06	.15	.18	−.13	.09
5.	.50	−.24	−.15	.11	.19	−.09	.13	−.12
6.	.31	−.24	.20	−.25	−.22	.13	.12	−.35
7.	.27	.18	−.36	−.42	.03	−.13	−.05	−.14
8.	.43	.11	−.26	−.45	.07	.19	−.12	−.21
9.	.40	.07	.04	.12	.13	−.14	.27	.11
10.	.56	−.45	−.15	.14	−.28	.21	−.15	−.08
11.	.70	.33	.20	−.04	.09	.14	−.18	−.09
12.	.50	−.25	−.20	.24	.17	−.05	.26	−.14
13.	.44	−.38	−.21	.11	−.10	−.03	−.04	.06
14.	.32	−.33	.05	−.35	.14	−.23	.05	−.06
15.	.35	.28	.22	.35	.20	.16	.14	.06
16.	.53	−.12	−.24	.28	.08	−.19	.13	.15
17.	.58	−.22	−.19	.19	.12	−.17	.17	.18
18.	.66	.07	.09	−.11	−.09	−.10	.10	−.34
19.	.30	−.36	.23	−.39	.30	−.29	−.22	.17
20.	.38	.17	−.28	−.15	−.08	.14	−.16	.24
21.	.35	−.50	.27	−.38	.28	−.28	−.31	.15
22.	.41	−.04	.08	−.09	−.34	−.10	−.25	.09
22a.	.51	.19	.36	−.03	−.05	−.03	−.05	.21
23.	.26	.08	−.22	.05	−.04	.15	.29	.46
24.	.38	.03	−.15	−.05	.17	.17	.13	.29
25.	.18	.44	−.38	.12	.17	.03	−.14	.07
26.	.20	.07	−.23	−.10	−.08	−.14	−.15	−.09
27.	.22	−.36	.14	−.27	−.20	.35	−.08	.24
27a.	.36	−.02	.03	−.07	.15	.23	.20	.05
28.	.38	−.10	.07	−.05	.18	.13	.14	−.09
29.	.22	−.10	.18	.46	.20	.11	.15	−.16
30.	.37	.31	−.47	.18	−.11	−.09	−.21	−.04
31.	.26	.33	−.34	−.30	−.17	.05	.20	−.02
32.	.42	.33	.25	.23	.23	.32	.13	.16
33.	.50	−.50	−.29	.27	−.25	.08	−.12	−.06
34.	.08	.11	−.31	.11	.40	.12	−.28	−.11
35.	.21	−.25	.18	−.27	−.21	.11	.27	.05
36.	.18	.39	−.29	−.34	−.09	−.13	.24	−.13
37.	.54	−.48	−.10	.03	−.26	.11	−.24	−.10
38.	.54	−.10	.06	.03	.13	−.06	−.20	−.15
39.	.58	−.07	.16	.09	−.06	.17	.07	.13
40.	.53	−.40	−.10	.17	−.04	−.06	−.03	.06

TABLE 2

Loadings of 42 Items from Psychiatric Interview Portion of MSRPP on Six Rotated Orthogonal Factors

	I	II	III	IV	V	VI
1.	.73	.10	.08	.15	.00	−.02
2.	.37	−.08	.13	−.13	.14	−.03
3.	.52	.15	.11	−.04	.02	.17
4.	.67	−.20	.03	.04	−.05	.02
5.	.10	.40	.12	.07	−.02	.45
6.	.07	.18	.43	.13	.30	−.08
7.	−.02	.02	.04	.63	−.07	.09
8.	.15	.13	.12	.63	−.12	.03
9.	.18	.07	.18	−.19	.33	.37
10.	.09	.77	.09	.08	.18	−.07
11.	.73	.18	.12	.32	−.02	.10
12.	.05	.45	−.03	−.04	.00	.52
13.	.03	.58	.12	.05	.07	.17
14.	−.03	.10	.50	.17	.17	.28
15.	.47	−.05	−.32	−.02	−.13	.23
16.	.10	.47	.10	.02	−.05	.52
17.	.13	.46	.12	.07	.02	.53
18.	.35	.28	.08	.28	.27	.29
19.	.05	.02	.80	.05	.00	.15
20.	.16	.25	.02	.48	−.10	−.03
21.	.02	.08	.87	.03	.08	.02
22.	.20	.27	.15	.22	.20	−.17
22a.	.55	.05	.22	.12	.23	.08
23.	.05	.15	−.23	.20	.11	.22
24.	.18	.17	.07	.23	−.03	.24
25.	.21	.03	−.27	.32	−.43	.14
26.	.00	.08	−.08	.28	.05	.22
27.	.05	.37	.34	.10	.30	−.22
27a.	.27	.13	.07	.15	.12	.22
28.	.30	.15	.15	−.03	.15	.22
29.	.32	.30	−.12	−.40	.02	.25
30.	.15	.32	−.28	.38	−.30	.10
31.	.02	.00	−.20	.62	.08	.12
32.	.67	.02	−.12	−.02	−.02	.20
33.	−.07	.82	.01	.00	.05	.03
34.	.11	.12	−.02	.10	−.56	.05
35.	−.01	.12	.19	.08	.52	.06
36.	−.07	.08	−.22	.57	.11	.30
37.	.07	.73	.28	.08	.12	−.07
38.	.33	.33	.25	.08	−.08	.15
39.	.47	.37	.07	.03	.24	.15
40.	.15	.52	.22	.00	.08	.24

Method

Each of the 120 patients in the study was rated by a team of two raters prior to the start of the study and after 3 and 6 months in the study. For each rating period, the two independent ratings were combined to yield measures with high reliability. Change scores were computed for each patient on each of the 42 rating-scale items by subtracting the 6 month rating from the initial rating. The change scores were intercorrelated after first being corrected to the separate drug group means. The correction procedure was introduced so that the change factors obtained in the subsequent analysis could be considered as representing primary dimensions of change in schizophrenic symptomatology and not primary dimensions of difference between treatment regimens. The centroid method of factor analysis was employed on the matrix of intercorrelations

TABLE 3

SYMPTOM CLUSTERS CORRESPONDING TO SIX ROTATED ORTHOGONAL FACTORS

Psychoticism
1. Irrelevance of responses to questions.
2. Inconsistency of thoughts and feelings.
3. Disorientation for time.
4. Elements of speech illogically and inconsistently connected.
5. Disorientation for person.
6. Speaks, mutters, mumbles to self.
7. Disorientation for place.
8. Perplexity.

Paranoidism
1. Tension.
2. Suspicion of influence.
3. Hostility, resentment, bitterness.
4. Unrealistic fears.
5. Anxiety over impulses to commit criminal or hostile act.
6. Apprehension.
7. Suspicion of persecution.
8. Suspicion of reference.

Guilt-conversion
1. Feelings of sinfulness and guilt.
2. Degree to which physical complaints not based upon organic factors.
3. Use of physical symptoms as defense mechanism.
4. Concern over health.

General retardation
1. Slowed motor reactions.
2. Reduced amount of speech.
3. Reduced loudness of speech.
4. Blocked speech.
5. Controlled, restrained, inhibited.
6. Reduced emotional responsiveness.

Depression
1. Feelings of inadequacy.
2. Depression.
3. Suicidal impulses.

Anxiety and tension
1. Tension.
2. Hostility, resentment, bitterness.
3. Anxiety over impulses to commit criminal or hostile act.
4. Apprehension.

between the change scores,[4] and rotation to an approximate orthogonal simple structure was accomplished by graphic methods.

RESULTS

The factor analysis resulted in a total of 11 initial unrotated centroid factors. The loadings of the 42 MSRPP variables on the 8 strongest factors are presented in Table 1, Factors 9 through 11 having only minimal loadings.

It was decided arbitrarily that variables with factor loadings of .40 or greater would be used in the interpretation of rotated factors.

[4] Computations were contracted for at the commercial rate from the Corporation for Economic and Industrial Research (CEIR), Arlington, Virginia.

That this high criterion might result in elimination of weak factors was realized; however, because the correlation matrix was estimated on only 117 degrees of freedom, it was considered desirable to restrict interpretation to only the major sources of common variance. Following rotation to an approximate orthogonal simple structure, six factors were found to have more than two loadings exceeding .40 in magnitude. The loadings of the 42 rating-scale items on the six rotated factors are presented in Table 2.

Table 3 identifies the content of the rating-scale items which were found to have loadings of .40 or higher on each of the six rotated orthogonal factors. The central concept which appears to represent the nature of the common variance in each set of items is included as the heading for each group of items. Fundamentally, of course, these symptom change concepts are defined in terms of the measures which correlate highly with them.

DISCUSSION

The first basic dimension of change in the psychiatric symptomatology of chronic schizophrenic patients appears to involve mental disorganization. Each symptom which changes with a change in this primary schizophrenic factor is pathognomonic of psychosis. None of the symptoms with high loadings on this factor have high loadings on any other factor obtained in this analysis. No other obtained factor contains symptoms which are pathognomonic of psychosis. Thus it is suggested that the mental disorganization factor represents the primary psychotic process in the chronic schizophrenic patients of this study. Factor I is independent of the mental processes represented by the other five factors, suggesting that the primary psychotic process can change without a corresponding change in the other unrelated processes.

It will be noted that a number of true psychotic symptoms, notably those concerned with hallucinatory behavior and motor disturbances, fail to have high loadings on any of the obtained change factors. If these symptoms were present in the patients, and if they had changed to any noticeable degree, then the items dealing with them would be expected to load on one of the obtained factors or to result in one or more additional factors.

There are several plausible explanations for the obtained results. The symptoms in question may have had minimal existence in the chronic patients; hence little change could be noted. Or these symptoms may be more stable and less subject to drug change than are the other psychotic symptoms. One explanation may be more reasonable for hallucinations and the other for mannerisms and posturing behavior.

In any event, only one psychoticism factor was obtained for the chronic patients involved in this study, and it must remain for future studies to determine whether other independent psychoticism factors exist for other patient populations. The finding of a unitary psychoticism factor appears to agree with conclusions advanced by Eysenck (1956) that there is a fundamental unitary process associated with the condition of being psychotic.

The second basic dimension of change in the psychiatric symptomatology of chronic schizophrenic patients appears to be related to distortion of thought accompanied by anxiety and tension. Although a delusional system may be involved, the process is one of distortion of the organized thought process rather than its disorganization. Since the distortion factor is independent of the disorganization factor, the results suggest that distortion of reality may occur at any level of mental organization; moreover, since change scores are being considered, change in degree of distortion need not be accompanied by change in degree of organization. Independent processes are apparently involved. The distortion factor is also independent of the other change factors resulting from the analysis.

The third basic dimension of change in psychiatric symptomatology in chronic schizophrenic patients appears to involve a guilt-conversion process. The factor is concerned with functional somatic complaints which accompany or which are accompanied by feelings of sinfulness and guilt. It is of some theoretical interest to note that changes in functional somatic symptoms correlate with changes in feelings of sinfulness and guilt, especially since these are the only psychological variables which correlate highly with the appearance and disappearance of physical symptoms. The fact that these symptoms change together suggests that a unitary process is involved. This factor appears to involve a neurotic process which is present in chronic schizophrenic patients but which is unrelated to the primary psychotic process.

The fourth independent dimension of change in the symptomatology of chronic schizophrenic patients is related to a general retardation process. Patients who are low on this factor become slower in movements, speak less, speak with reduced volume, and evince reduced emotional responsiveness. The fact that these several symptoms change together suggests that they are dependent upon a common underlying process which is possibly related to vital energy level. This factor has been found to be especially sensitive to drug effects, with some ataractic drugs given at high doses producing a marked increase in general retardation.

The fifth basic dimension of change in the psychiatric symptomatology of chronic schizophrenic patients appears to represent the mood component of the affective process. Changes in mood, self-respect, and the occurrence of suicidal impulses seem to be related to a unitary mental process. It is interesting that depression is a psychiatric construct which is usually conceived to involve changes in mood, retardation, and somatic symptomatology; however, in the present analysis these three components were found to change independently of one another.

The sixth basic dimension of symptom change in chronic schizophrenic patients is somewhat more difficult to interpret than the other five because it involves only that part of anxiety and tension which is independent of distortion in thought processes.[5] An increase in distortion of reality is accompanied by a corresponding increase in anxiety and tension, but anxiety and tension may increase independently of distortion in thought processes. As with other nonpsychotic primary processes, anxiety and tension appear as neurotic elements in psychotic patients, and the symptoms involved may change without a corresponding change in primary psychoticism.

[5] A final oblique rotation would have yielded an "anxiety" factor which was substantially correlated with the "paranoidism" factor.

Factor analysis leads to a conception of the psychiatric patient as characterized by a limited number of functionally independent processes which are reflected in observable symptomatology. Each patient is conceptualized as having a level of functioning related to each of the primary processes. It is interesting to note that the primary dimensions of change obtained from the present analysis are adequate in describing several "types" of patients included in the nomenclature of the American Psychiatric Association Mental Hospital Service (1952) classification system. Since the disorganization factor represents the only truly psychotic factor evident in these patients, it is conceived as the core of the schizophrenic process. A pure type of schizophrenic may evince pathology in this area only. The psychotic process may be superimposed on a personality pattern reflecting other independent processes. In the case where the disorganization process runs concurrently with a marked distortion process, "schizophrenic reaction, paranoid type," is the result. The present results suggest that two independent processes, not one, are responsible for this classification. Patients may show a marked distortion process without accompanying disorganization. In such a case, "paranoid reaction" would be the classification.

No special nomenclature is available to describe the schizophrenic patient with a primary neurotic process centering about guilt and somatic complaints. Hypochondriasis, functional somatic disorders, and somatic delusions are common early in several "types" of schizophrenic reactions. The important thing to note here is that over a reasonably short period of only 6-months duration, change in this neurotic process is uncorrelated with change in the primary schizophrenic process.

The general retardation process, when extreme and in association with the mental disorganization process, may result in the classification of "schizophrenic reaction, catatonic type." If the retardation is less severe and is accompanied by a marked depression of mood, the picture is one of "schizophrenic reaction, schizoaffective type." In cases where the mood factor predominates and in cases where mood, retardation, and somatic processes coexist, the classification may be one of "depressive reaction." Without the point being labored further, one can recognize how other neurotic and psychotic reaction types can be conceived to result from various combinations of the independent processes. The most interesting thing is that such a comprehensive set of processes can be identified from change in the symptomatology of chronic schizophrenic patients and that the several processes are independent over the period of time considered in this study.

The fact that the several processes appear to be independent over a 6-month observation period does not establish their independence in the long-term. Risking pure conjecture for the purpose of elaborating this point, one may conceive the schizophrenic process as being psychogenic in origin and even as resulting directly from one or more of the observed neurotic processes. This would not require the processes to have a short-term relationship. They could operate in different media. On the other hand, the schizophrenic process may be conceived as an entirely separate disease process which superimposes itself upon an existing set of personality and adjustment processes. The several processes may appear to be related because disorganization renders the individual less capable of coping with the world and hence brings the neurotic processes into the foreground. It may be only in stabilized, chronic psychotic patients that we are able to observe the fact that several independent processes are involved.

As a final point to be considered in this discussion, it is interesting to examine the similarities and differences between the change factors obtained in this analysis and the status factors obtained by Lorr (1953). The number of first order change factors identified in this analysis was smaller than the number of factors identified by Lorr in analysis of status scores. Lorr's factors include ward behavior items in addition to those from the psychiatric interview; his sample was considerably more heterogeneous, and he worked with "between-patient" variance rather than "within-patient-over-time" variance. Nevertheless, some general agreement is present. The disorganization factor in the present analysis

corresponds closely to the "conceptual disorganization" factor identified by Lorr. In the present analysis, all three of the "orientation" items loaded heavily on this single factor, whereas only "orientation for time" was included in the corresponding MSRPP factor. The MSRPP factor also includes "false beliefs," which would seem more appropriate in a "distortion" than in a "disorganization" factor.

The distortion factor obtained in the present analysis includes the scales subsumed under Lorr's (1953) "paranoid projection" factor. In this case, the MSRPP factor derived from between-patient variance appears to be a conceptually purer factor than the distortion factor derived from the change scores. Apparently, paranoid projection can appear in individuals with any characteristic level of anxiety (lack of between-patient covariance), but an increase in distortion results in an increase in the characteristic anxiety level which is present.

The guilt-conversion factor in the present analysis appears to correspond directly to Lorr's (1953) "hysteric conversion" factor, which he has marked as "tentative," with the exception that the guilt item does not appear in the MSRPP factor.

The general retardation factor from the present study includes items from "retarded depression," "activity," "melancholy agitation," and "conceptual disorganization" factors of the MSRPP. There is considerable overlap among the first three of these MSRPP factors with several items appearing on more than one of them. It appears reasonable that a single concept should be involved in all three. Similarly, the mood factor in the present analysis appears to correspond to another common source of variance included in the "retarded depression," "activity level," and "melancholy agitation" factors from the MSRPP. It is concluded here that analysis of change scores in these areas provides more meaningful and unitary constructs than were produced by analysis of status scores.

No pure anxiety factor was obtained in the previous analysis based upon status scores. The "melancholy agitation" factor obtained by Lorr (1953) contains anxiety and tension items, but the present results lead to the suggestion that anxiety and tension represent a useful independent construct of individual difference in psychiatric patients.

The symptom clusters identified in the present analysis are homogeneous and readily interpretable in terms of unitary underlying constructs. The method of examining psychiatric processes through analysis of change scores is recommended. It is recognized that this work represents only the initial step in the establishing of a reliable and comprehensive description of basic psychiatric processes which could result from analysis of the covariation of symptom change. The utility of objective psychiatric rating scales, such as the one developed by Lorr (1953), is strongly supported by the results.

SUMMARY

Data from a previously reported chemotherapy study involving 120 chronic schizophrenic patients were reanalyzed for the purpose of identifying primary dimensions of change in psychiatric symptomatology. Six independent processes were identified through a centroid factor analysis and orthogonal rotation. The processes are identified with mental disorganization, distortion of thought processes, guilt-conversion, retardation, depression in mood, and anxiety. The several processes are independent of one another over a reasonably short period of time in chronic schizophrenic patients. Only the disorganization process involves unquestionably psychotic symptomatology; it appears to be identified with the primary schizophrenic process. Various subtypes of schizophrenic patients can be described in terms of this psychotic process being superimposed upon a personality marked by emphasis of one or more of the other independent processes.

REFERENCES

AMERICAN PSYCHIATRIC ASSOCIATION, Mental Hospital Service, Committee on Nomenclature and Statistics of the American Psychiatric Association. *Diagnostic and statistical manual: Mental disorders.* Washington, D. C.: APA, 1952.

EYSENCK, S. B. G. Neurosis and psychosis: An experimental analysis. *J. ment. Sci.*, 1956, **102**, 517–529.

LORR, M. *Multidimensional scale for rating psychiatric patients.* Technical Bulletin TB10–507, 1953, Veterans Administration, Washington, D. C.

SHAWVER, J. R., GORHAM, D. R., LESKIN, L. W., GOOD, W. W., & KABNICK, D. E. Comparison of chlorpromazine and reserpine in maintenance drug therapy. *Dis. nerv. System*, 1959, **20**, 452–457.

(Received December 17, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 63, No. 3, 603–606

PREPARATORY ACTION AND BELIEF IN THE PROBABLE OCCURRENCE OF FUTURE EVENTS[1]

RUBY B. YARYAN AND LEON FESTINGER
Stanford University

PEOPLE are frequently faced with the possibility that a future event that would have important consequences for them may actually occur. Usually, in such a situation, it is possible to take some action in preparation for the possible future event so that, if it should occur, its impact on the person would be more favorable than would otherwise have been the case. Such preparatory action, however, may involve considerable inconvenience and, if the possible future event should not occur, the preparatory action will have been useless.

Obviously, the more a person in such a situation believes that the possible future event will really occur, the more willing he should be to take action in preparation for it. This statement is almost self-evident and, of course, is *not* our concern here. The problem we wish to deal with is, indeed, the reverse of this. That is, if a person has somehow been induced to take preparatory action of a rather inconvenient or effortful kind, will this, in and of itself, lead him to believe that the future event is even more likely to occur?

The theory of dissonance (Festinger, 1957) would lead one to give an affirmative answer to this question. Consider the cognitions of a person who has expended considerable effort preparing for some possible future event and who also knows many things that indicate to him that the event will not actually occur. These two sets of cognitions are dissonant with one another. Considering only those facts the person knows which indicate that the event will not occur, it would follow psychologically that the person would not expend effort in preparation for it. Hence, the information that he did expend effort in preparation is dissonant with information he has that indicates the event will not occur.

Given this dissonance, then, let us examine how it may be reduced. There are three major ways in which such dissonance can be reduced:

1. The person may somehow be able to persuade himself that the preparation was really effortless and not at all inconvenient.

2. The person may find for himself some other objectives, in addition to the possible future event, which "justify" the actions he has taken.

3. The person may persuade himself that the possible future event is, indeed, very likely to occur.

To the extent that this last means of reducing dissonance is employed by people, we emerge with the prediction that, after a person engages in effortful preparation for a future event, he should believe the event is even more likely to occur than he had previously believed. The present experiment was designed to test this derivation under controlled laboratory conditions.

METHOD

The basic design of the experiment was simple. All subjects were given a standard set of instructions concerning the exact probability of occurrence of a future event (that they might have to take an examination). One group of subjects, selected at random, was then induced to engage in effortful preparation while another group was induced to engage in relatively effortless preparation. The major dependent variable in which we were interested was how likely the subject thought it was that the event would actually occur (that she herself would have to take the examination). The details of the procedure follow.

Subjects and Volunteering Procedure

The subjects were 60 female high school students, ranging from 14 to 18 years of age, who had volunteered to participate in an experiment on "Techniques of Studying."

In asking for volunteers the experimenter said:

We are doing research on the techniques of studying; specifically, we are looking into the guessing process that seems to be involved when students study for tests.

Now, here is what we are asking you to do. We are requesting that everyone volunteer for two hours. The experiment takes one hour for 50% of the subjects and two hours for the other 50%. I can't tell you now in which group, the one hour or the two hour group, you will be placed, for this is part of the

[1] The research reported in this paper was supported by National Science Foundation Grant G-11255 to the second author.

experiment itself. That is why we are asking everyone to volunteer for two hours, although half of you will be finished at the end of one hour.

Thus, before the experiment each subject knew that, even though each appointment was for 2 hours, 50% of the subjects would be finished at the end of 1 hour.

Course of the Experiment

When each subject arrived for the experiment she was told the following:

We are studying the techniques that students use to study for exams. We know that in the process of studying, students form hypotheses, hunches, or guesses about an exam (for example, if there will be essays on the exam, what the essays will be about, the type of material to be asked, and so on), and then they orient their behavior according to their guesses. As yet we don't know the hows or the whys or the types of conclusions that students reach and their relative accuracy. These are the questions we hope to answer in this experiment.

As I told you before, this experiment takes one hour for half of the subjects and two hours for the others. We won't tell you which group you are in until we finish the first part of the experiment. We have the experiment designed this way so that we can get at the answers we want. What we have done is to divide the people that have volunteered for the experiment into two experimental conditions; we have done this on the basis of certain criteria (we have access to high school records and other things). Now I can't tell you what these criteria are at this point, for we want you to guess about them later. In short, however, you have already been put in one of the two conditions, so no matter what you do now this will not change.

In one condition the people will take a new, highly reliable aptitude test, one developed specifically for people above the ages of 14. If you happen to be in this condition, the time you will be here will be two hours. The other half of the people will not take any aptitude test at all; these people will be finished in one hour. We had everyone volunteer for two hours, so that people wouldn't know which condition they were in.

So that you will be familiar with the total situation, even though you may or may not have to take the test, we tell everyone generally about it.

As I mentioned before, this test that half the people that signed up will be taking is a new type of aptitude or IQ test, "The Revised Standardized Aptitude Test 1958." It has been specifically developed for people above the ages of 14 and has proved to be very successful in accurately indicating IQ. The test goes on the assumption that the reason past IQ tests have been unreliable with people above these ages is that through studying specific areas, these students have a specialized knowledge of the factual information in the area of their concentration and relatively little extensive knowledge of the information in other areas. This, you can see, leads to noncomparable groups of people and, as a consequence, leads to noncomparable measures of the aptitudes. In order to get around this problem and to standardize the conditions for all students, this test requires that the people taking it be given some specific information beforehand, which will give them all the necessary tools to work with the questions on the test; this way everyone approaches the test with the same information. What the test measures is not this information, but rather the way that they work with the information that they have been given.

In order to establish the "high preparatory effort condition," the experimenter said:

Now in a few minutes I shall give you the "Information Sheet." As you will see, it is crucial for taking the test. If you hapen to be one of those selected to take it, which you may not be, it is of utmost importance that you memorize these definitions, for the test is not valid if the people do not learn them. We are requiring everyone to commit these definitions to memory, no matter which condition they are in.

Thus far, it has taken people around 25–30 minutes to memorize these. You may take as much time as you need to learn this material thoroughly, for you will not be able to have this sheet later, if you happen to be one of those previously designated to take the test.

After you have memorized these definitions, we will give you a short questionnaire and interview; then we will tell you in which group you have been placed. If you have not been selected to take the test, you can leave then; if you are one of the people that have been selected, you will then take the test.

The "low preparatory effort condition" was established by the following instructions:

Now in a few minutes I shall give you the "Information Sheet." As you will see, it is crucial for taking the test, if you happen to be one of those selected to take it, which you may not be. In any case, we want everyone to briefly look over it.

You will have a minute or so to look over the information, just to get an idea of what it is all about. You need not attempt to commit it to memory or study it, for you will be able to have the "Information Sheet" with you, if you should be one of the people previously designated to take the test.

After you have looked over these definitions, we will give you a short questionnaire and interview; then we will tell you in which group you have been placed. If you have not been selected to take the test, you can leave then; if you are one of the people that have been selected, you will then take the test.

Each subject was then handed an "Information Sheet" containing 25 symbols with arbitrary abstract definitions of each symbol. After each subject had finished either memorizing the list (high effort condition) or briefly familiarizing herself with it (low effort condition), she was asked a number of questions which we will describe exactly when we discuss the results of the experiment. When the questionnaire had been completed, the true nature of the experiment was explained to each subject in detail. Of course, no aptitude test was administered.

Although each subject was asked not to talk about the experiment with others in the school, some talking about the details of the experiment did occur. Seven subjects, four from the "high effort" and three from

the "low effort" condition, had to be discarded because they had heard what the experiment was all about. Such subjects, of course, realized beforehand that there was no aptitude test to take. Two other subjects were discarded (one from each condition) because of misunderstanding of instructions and directions. Valid data remained for 51 subjects, 26 in the high preparatory effort condition and 25 in the low preparatory effort condition.

RESULTS

Effectiveness of the Manipulation

In order to determine whether or not we had effectively created a high preparatory effort and a low preparatory effort condition, we asked all subjects the following question:

Do you think that the work you had to do in preparation was . . .

(−3) very easy	(−2) somewhat easy	(−1) slightly easy	(+1) slightly difficult	(+2) somewhat difficult	(+3) very difficult

The mean rating for subjects in the high preparatory effort condition was +.7. The comparable figure for those in the low preparatory effort condition was −.9. The difference between the two conditions is significant beyond the .05 level. In other words, our experimental manipulation was successful. Those in the high effort condition felt they had worked harder in preparation for the test than did those in the low effort condition.

Expectation of Taking the Test

Let us recall the hypothesis this experiment was designed to test. We would expect, from considering the theory of dissonance, that, if a person is induced to engage in effortful activity in preparation for a possible future event, the person will subsequently believe that the event is more likely to occur. If this hypothesis is correct, we would expect the subjects in the high preparatory effort condition to believe that they would actually have to take the test more than would subjects in the low preparatory effort condition. To obtain the data necessary to test this hypothesis, we asked the subjects the following question:

Do you believe that you are one of the people that have been selected to take the test?

(+3) certain	(+2) probably	(+1) slightly likely	(−1) slightly unlikely	(−2) probably not	(−3) definitely not

The mean rating on this scale for subjects in the high effort condition was +1.1; for those in the low effort condition the mean rating was only +.2. The difference is significant at the .01 level ($t = 2.70$). Another way of looking at these data is in terms of the number of subjects who checked the plus side of the scale, indicating that they thought they would probably have to take the test. In the high effort condition 24 of the 26 subjects (92%) checked the plus side of the scale while in the low effort condition only 15 out of 25 (60%) checked this side of the scale. In other words, as a group, the subjects in the low effort condition reflect fairly well what they have been told, namely, that 50% of them would have to take the test. The low effort condition subjects divided themselves 60% to 40%, which is not very different from an "expected" 50%. In the high effort condition, however, there is clearly a stronger belief that they would have to take the test. Since subjects were allocated to experimental conditions at random, it seems clear that this stronger belief in having to take the test is a consequence of the amount of effort expended in preparation. We may confidently conclude that the data support the hypothesis.

Other Measures

In the post-experimental questionnaire, three other questions were asked in an attempt to obtain control data. That is, these questions were asked in order to show that the difference we observed in how likely subjects thought it was that they would have to take the test was a specific difference related to dissonance reduction. Let us explain a bit more what we mean by this. The notion behind the study was that there is dissonance between knowing one has expended considerable effort in preparation for a test and knowing that the test is not very likely to take place. Persuading oneself that the test is, indeed, likely to take place obviously reduces this dissonance. But we would clearly be more confident that the obtained difference was due to dissonance reduction if we could show that on other measures, which were related to the situation but served no dissonance reducing function, there were no differences between the two groups.

The three questions we used were chosen

to be closely related to the testing situation which the subjects presumably faced. On none of the questions was any difference between the two experimental conditions apparent. One question concerned how important they thought IQ measures were, another concerned the perceived difficulty of the possible test, and the third asked them how anxious they were about possibly having to take the test. The means for the two experimental groups are very close together on each question. On importance of IQ tests the means are +1.38 and +1.40; on perceived difficulty the means are +1.08 and +.96; and for anxiety about the test the means are −.23 and −.40. All the differences are clearly negligible. We may then conclude that the difference in belief that they would have to take the test is not a general difference between the conditions but is specific to dissonance reduction.

An Incidental Finding

A recent study by Schachter (1959) reports a relationship between birth order and experienced anxiety when subjects are faced with a threatening situation. Specifically, Schachter found that when subjects were told they were to receive a strong electric shock, those who were first born or who were only children experienced more anxiety than did those who were later born.

Although this whole issue is not of direct relevance to the present experiment, we took the opportunity to verify Schachter's (1959) finding in a slightly different context. While Schachter used electric shock to produce an anxiety arousing situation, his finding should hold more generally if it is indeed valid. The situation in our experiment may also be seen as a threatening and anxiety arousing situation. After all, the possibility of taking a test does arouse anxiety for many students. It is, of course, easy to ask subjects whether they are only children, first born or later born, and so we collected these data.

If Schachter's (1959) finding is valid we would expect birth order to be related to the responses the subjects gave on the question that asked them how anxious they were about the prospect of having to take the examination. The subjects answered the question on a scale from −3 "feel very comfortable about it," to +3 "feel very uncomfortable about it." Throwing both experimental conditions together (there were no differences at all between the two experimental conditions), we find that the 32 subjects who were first born or only children show a mean of −.03, just about neutral on the scale. There were 18 later born subjects and these give a mean of −.98, considerably in the direction of saying they feel comfortable about it. This difference is significant between the .05 and .10 levels ($t = 1.7$). In other words, later born subjects do tend to have less experienced anxiety than first born or only children. Although the difference is not strongly significant, it provides clear support for Schachter's findings. It may also be pointed out that birth order was found to relate to anxiety only and to nothing else that we measured in this experiment.

SUMMARY

An experiment was conducted to test the following hypothesis derived from Festinger's (1957) theory of cognitive dissonance: If people exert a great deal of effort (memorizing a list of symbolic definitions) in preparation for a possible future event (an intelligence test), they should believe more strongly in the likelihood of occurrence of the event (that they would have to take the test) than people who do not have to engage in effortful preparatory activity. The results of the experiment provide strong support for the hypothesis.

REFERENCES

FESTINGER, L. *A theory of cognitive dissonance.* Evanston, Ill.: Row, Peterson, 1957.

SCHACHTER, S. *The psychology of affiliation.* Stanford: Stanford Univer. Press, 1959.

(Received December 18, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 63, No. 3, 607-618

PSYCHIATRIC DIAGNOSIS:

A CRITIQUE[1]

EDWARD ZIGLER
Yale University

AND

LESLIE PHILLIPS
Worcester State Hospital and Clark University

THE inadequacies of conventional psychiatric diagnosis have frequently been noted (Ash, 1949; Cattell, 1957; Eysenck, 1952; Foulds, 1955; Harrower, 1950; Hoch & Zubin, 1953; Jellinek, 1939; King, 1954; Leary & Coffey, 1955; Mehlman, 1952; Menninger, 1955; Noyes, 1953; Phillips & Rabinovitch, 1958; Roe, 1949; Rogers, 1951; Rotter, 1954; Scott, 1958; Thorne, 1953; Wittenborn & Weiss, 1952; Wittman & Sheldon, 1948). The responses to this rather imposing body of criticism have ranged from the position that the present classificatory system is in need of further refinement (Caveny, Wittson, Hunt, & Herman, 1955; Foulds, 1955), through steps towards major revisions (Cattell, 1957; Eysenck, 1952; Leary & Coffey, 1955; Phillips & Rabinovitch, 1958; Thorne, 1953; Wittman & Sheldon, 1948), to a plea for the abolishment of all "labeling" (Menninger, 1955; Noyes, 1953; Rogers, 1951). As other investigators have noted (Caveny et al., 1955; Jellinek, 1939), this last position suggests that the classificatory enterprise is valueless. This reaction against classification has gained considerable popularity in clinical circles. The alacrity with which many clinicians have accepted this view seems to represent more than a disillusionment with the specific current form of psychiatric diagnosis. These negative attitudes appear to reflect a belief that diagnostic classification is inherently antithetical to such clinically favored concepts as "dynamic," "idiographic," etc. Thus, a question is raised as to whether any diagnostic schema can be of value. Let us initially direct our attention to this question.

On Classification

The growth among clinicians of sentiment against categorization has coincided with a period of critical reappraisal within the behavioral sciences generally (Beach, 1950; Brower, 1949; Cronbach, 1957; Guthrie, 1950; Harlow, 1953; Koch, 1951; MacKinnon, 1953; Marquis, 1948; Rapaport, 1947; Roby, 1959; Scott, 1955; Tolman, 1953; Tyler, 1959). This parallel development is more than coincidental. The reaction against "labeling" can be viewed as an extreme outgrowth of this critical self-evaluation, i.e., that psychology's conceptual schemata are artificial in their construction, sterile in terms of their practical predictions, and lead only to greater and greater precision about matters which are more and more irrelevant. It is little wonder that in this atmosphere, conceptualization has itself become suspect nor that Maslow's (1948) exposition of the possible dangers of labeling or naming has been extended (Rotter, 1954) as a blanket indictment of the categorizing process.

The error in this extension is the failure to realize that what has been criticized is not the conceptual process but only certain of its products. The criticisms mentioned above have not been in favor of the abolishment of conceptualization, but have rather been directed at the prematurity and rarifications of many of our conceptual schemata and our slavish adherence to them. Indeed, many of these criticisms have been accompanied by pleas for lower-order conceptualization based more firmly on observational data (Koch, 1951; MacKinnon, 1953; Tolman, 1953).

In the clinical area, the sentiment against classification has become sufficiently serious that several investigators (Cattell, 1957; Caveny et al., 1955; Eysenck, 1952; Jellinek, 1939) have felt the need to champion the merits of psychiatric categorization. They have pointed out that diagnosis is a basic scientific classificatory enterprise to be viewed as essentially the practice of taxonomy, which is characteristic of all science. Eysenck (1952) puts the matter quite succinctly in his statement, "Measurement is essential to

[1] This investigation was supported by the Dementia Praecox Research Project, Worcester State Hospital, and a research grant (M-896) from the National Institute of Mental Health, United States Public Health Service.

science, but before we can measure, we must know what it is we want to measure. Qualitative or taxonomic discovery must precede quantitative measurement" (p. 34).

Reduced to its essentials, diagnostic classification involves the establishment of categories to which phenomena can be ordered. The number of class systems that potentially may be constructed is limited only by man's ability to abstract from his experience. The principles employed to construct such classes may be inductive, deductive, or a combination of both, and may vary on a continuum from the closely descriptive to the highly abstract.

Related to the nature of the classificatory principle are the implications to be derived from class membership. Class membership may involve nothing more than descriptive compartmentalization, its only utility being greater ease in the handling of data. Obversely, the attributes or correlates of class membership may be widespread and far-reaching in their consequences. The originators of a classificatory schema may assert that specified behavioral correlates accompany class membership. This assertion is open to test. If the hypothesized correlates represent the full heuristic value of the diagnostic schema and class membership is found not to be related to these correlates, then revision or discard is in order. A somewhat different type of problem may also arise. With the passage of time, correlates not originally related to the schema may erroneously be attributed to class membership. Nevertheless, the original taxonomy may still possess a degree of relevance to current objectives in a discipline. In these circumstances, its maintenance may be the rational choice, although a clarification and purification of categories is called for. The relationship of the two problems outlined here to the criticism of contemporary psychiatric diagnosis will be discussed later. What should be noted at this point is that the solution to neither problem implies the abolishment of the attempt at classification.

Another aspect of taxonomy is in need of clarification. When a phenomenon is assigned to a class, certain individual characteristics of that phenomenon are forever lost. No two class members are completely identical. Indeed, a single class member may be viewed as continuously differing from itself over time. It is this loss of uniqueness and an implied unconcern with process that have led many clinicians to reject classification in principle. While classificatory schemata inevitably involve losses of this type, it must be noted that they potentially offer a more than compensatory gain. This gain is represented in the significance of the class attributes and correlates. Class membership conveys information ranging from the descriptive similarity of two phenomena to a knowledge of the common operative processes underlying the phenomena.

A conceptual system minimizes the aforementioned loss to the extent that only irrelevant aspects of a phenomenon are deleted in the classificatory process. The implicit assumption is made that what is not class relevant is inconsequential. The dilemma, of course, lies in our lacking divine revelation as to what constitutes inconsequentiality. It is this issue which lies at the heart of the idiographic versus nomothetic controversy (Allport, 1937, 1946; Beck, 1953; Eysenck, 1954; Falk, 1956; Hunt, 1951a, 1951b; Skaggs, 1945, 1947). The supporters of the idiographic position (Allport, 1937; Beck, 1953) have criticized certain conceptual schemata for treating idiosyncratic aspects of behavior as inconsequential when they are in fact pertinent data which must be utilized if a comprehensive and adequate view of human behavior is to emerge. However, the idiographic position is not a movement toward the abolishment of classification, a fact emphasized by Allport (1937) and Falk (1956). Rather, it represents a plea for broader and more meaningful classificatory schemata.

A conceptually different type of argument against the use of any diagnostic classification has been made by the adherents of nondirective psychotherapy (Patterson, 1948; Rogers, 1946, 1951). This position has advanced the specific contention that differential diagnosis is unnecessary for, and perhaps detrimental to, successful psychotherapy. This attitude of the nondirectivists has been interpreted (Thorne, 1953) as an attack on the entire classificatory enterprise. To argue against diagnosis on the grounds that it affects therapeutic outcome is to confuse diagnosis as an act of scientific classification with the

present clinical practice of diagnosis with its use of interviewing, psychological testing, etc. The error here lies in turning one's attention away from diagnosis as an act of classification, a basic scientific enterprise, and attending instead to the immediate and prognostic consequences of some specific diagnostic technique in a specific therapeutic situation, i.e., an applied aspect. To reject the former on the basis of the latter would appear to be an unsound decision.

Although the nondirectivists' opposition to diagnosis seems to be based on a confusion between the basic and applied aspects of classification, implicitly contained within their position is a more fundamental argument against the classificatory effort. Undoubtedly, diagnosis both articulates and restricts the range of assumptions which may be entertained about a client. However, the philosophy of the nondirectivist forces him to reject any theoretical position which violates a belief in the unlimited psychological growth of the client. It would appear that this position represents the rejection, in principle, of the view that any individual can be like another in his essential characteristics, or that any predictable relationship can be established between a client's current level of functioning and the ends which may be achieved. In the setting of this assumption, a transindividual classificatory schema is inappropriate. There is no appeal from such a judgment, but one should be cognizant that it rejects the essence of a scientific discipline. If one insists on operating within the context of a predictive psychology, one argues for the necessity of a classificatory system, even though particular diagnostic schemata may be rejected as irrelevant, futile, or obscure.

Let us now direct our discussion toward some of the specific criticisms of conventional psychiatric diagnosis—that the categories employed lack homogeneity, reliability, and validity.

Homogeneity

A criticism often leveled against the contemporary diagnostic system is that its categories encompass heterogeneous groups of individuals, i.e., individuals varying in respect to symptomatology, test scores, prognosis, etc. (King, 1954; Rotter, 1954: Wittenborn, 1952; Wittenborn & Bailey, 1952; Wittenborn & Weiss, 1952). Contrary to the view of one investigator (Rotter, 1954), a lack of homogeneity does not necessarily imply a lack of reliability. King (1954) has clearly noted the distinction between these two concepts. Reliability refers to the agreement in assigning individuals to different diagnostic categories, whereas homogeneity refers to the diversity of behavior subsumed within categories. While the two concepts may be related, it is not difficult to conceptualize categories which, though quite reliable, subsume diverse phenomena.

King (1954) has argued in favor of constructing a new diagnostic classification having more restrictive and homogeneous categories. He supports his argument by noting his own findings and those of Kantor, Wallner, and Winder (1953), which have indicated that within the schizophrenic group subcategories may be formed which differ in test performance. King found further support for the construction of new and more homogeneous diagnostic categories in a study by Windle and Hamwi (1953). This study indicated that two subgroups could be constructed within a psychotic population which was composed of patients with diverse psychiatric diagnoses. Though matched on the distribution of these diagnostic types, the subgroups differed in the relationship obtained between test performance and prognosis. On the basis of these studies, King suggests that the type of homogeneous categories he would favor involves such classificatory dichotomies as reactive versus process schizophrenics and chronic versus nonchronic psychotics.

An analysis of King's (1954) criticism of the present diagnostic system discloses certain difficulties. The first is that King's heterogeneity criticism does not fully take into consideration certain basic aspects of classification. A common feature of classificatory systems is that they utilize classes which contain subclasses. An example drawn from biology would be a genus embracing a number of species. If schizophrenia is conceptualized as a genus, it cannot be criticized on the grounds that all its members do not share a particular attribute. Such a criticism would involve a confusion between the more specific

attributes of the species and the more general attributes of the genus. This is not to assert that schizophrenia does in fact possess the characteristics of a genus. It is, of course, possible that a careful analysis will reveal that it does not, and the class schizophrenia will have to be replaced by an aggregate of entitites which does constitute a legitimate genus. However, when a genus is formulated, it cannot be attacked because of its heterogeneous nature since genera are characterized by such heterogeneity.

A more serious difficulty with King's (1954) heterogeneity criticism lies in the inherent ambiguity of a homogeneity-heterogeneity parameter. To criticize a classificatory system because its categories subsume heterogeneous phenomena is to make the error of assuming that homogeneity is a quality which inheres in phenomena when in actuality it is a construction of the observer or classifier. In order to make this point clear, let us return to King's argument. What does it mean to assert that chronic psychosis is an example of an homogeneous class, while schizophrenia is an example of an heterogeneous one? In terms of the descriptively diverse phenomena encompassed, the latter would appear to have the greater homogeneity. The statement only has meaning insofar as a particular correlate—for instance, the relationship of test score to prognosis—is shared by all members of one class but not so shared by the members of the other class. Thus, the meaningfulness of the homogeneity concept is ultimately dependent on the correlates or attributes of class membership or to the classificatory principle related to these correlates or attributes. The intimacy of the relationship between the attributes of classes and the classificatory principle can best be exemplified by the extreme case in which a class has but a single attribute, and that attribute is defined by the classificatory principle, e.g., the classification of plants on the basis of the number of stamens they possess. Therefore, the heterogeneity criticism of a classificatory system is nothing more than a plea for the utilization of a new classificatory principle so that attention may be focused on particular class correlates or attributes not considered in the original schema. While this plea may be a justifiable one, depending on the significance of the new attributes, it has little to do with the homogeneity, in an absolute sense, of phenomena. Indeed, following the formulation of a new classificatory schema, the heterogeneity criticism could well be leveled against it by the adherents of the old system, since the phenomena encompassed by the new categories would probably not be considered homogeneous when evaluated by the older classificatory principle.

Although differing in its formulation, the heterogeneity criticism of present psychiatric classification made by Wittenborn and his colleagues (Wittenborn, 1952; Wittenborn & Bailey, 1952; Wittenborn & Weiss, 1952) suffers from the same difficulties as does King's (1954) criticism. Wittenborn's findings indicated that individuals given a common diagnosis showed differences in their symptom cluster score profiles based on nine symptom clusters isolated earlier by means of factor analytic techniques (Wittenborn, 1951; Wittenborn & Holzberg, 1951). It is upon the existence of these different profiles within a diagnostic category that Wittenborn bases his heterogeneity criticism. Here again the homogeneity-hetorogeneity distinction is only meaningful in terms of an independent criterion, a particular symptom cluster score profile. Had it been discovered that all individuals placed into a particular diagnostic category shared a common symptom cluster score profile, then this category would be described as subsuming homogeneous phenomena. But the phenomena—the symptoms mirrored by the symptom profile—are not homogeneous in any absolute sense because the pattern of symptoms may involve the symptoms in descriptively diverse symptom clusters. Thus, the homogeneity ascribed to the category would refer only to the fact that individuals within the category homogeneously exhibited a particular pattern of descriptively diverse behaviors. However, the organization of symptoms mirrored by the symptom cluster profiles is not in any fundamental sense different from that observed in conventional diagnostic syndromes. Both methods of categorization systematize diverse behaviors because of an observed regularity in their concurrent appearance.

The difference between these two approaches, then, lies only in the pattern of deviant behaviors that define the categories. Indeed, Eysenck (1953) has noted that both the clinician and the factor analyst derive syndromes in essentially the same manner, i.e., in terms of the observed intercorrelations of various symptoms. It is the difference in method, purely observational versus statistical, that explains why the final symptom structure may differ. The assumption must not be made that the advantage lies entirely with the factor analytic method. The merit accruing through the greater rigor of factor analysis may be outweighed by the limitations imposed in employing a restricted group of symptoms and a particular sample of patients. Thus, the factor analyst cannot claim that the class-defining symptom pattern he has derived is a standard of homogeneity against which classes within another schema can be evaluated. The plea that symptom cluster scores, derived from factor analytic techniques, substitute for the present method of psychiatric classification has little relevance to the heterogeneity issue.

In the light of this discussion we may conclude that the concept of homogeneity has little utility in evaluating classificatory schemata. Since the heterogeneity criticism invariably involves an implicit preference for one classificatory principle over another, it would perhaps be more fruitful to dispense entirely with the homogeneity-heterogeneity distinction, thus, allowing us to direct our attention to the underlying problem of the relative merits of different classificatory principles.

Reliability and Validity

A matter of continuing concern has been the degree of reliability of the present diagnostic system. Considerable energy has been expended by both those who criticize the present system for its lack of reliability (Ash, 1949; Boisen, 1938; Eysenck, 1952; Mehlman, 1952; Roe, 1949; Rotter, 1954; Scott, 1958) and those who defend it against this criticism (Foulds, 1955; Hunt, Wittson, & Hunt, 1953; Schmidt & Fonda, 1956; Seeman, 1953). Certain investigators (Foulds, 1955; Schmidt & Fonda, 1956) who have offered evidence that the present system is reliable have also pointed out that the earlier studies emphasizing the unreliability of psychiatric diagnosis have suffered from serious conceptual and methodological difficulties.

In evaluating the body of studies concerned with the reliability of psychiatric diagnosis, one must conclude that so long as diagnosis is confined to broad diagnostic categories, it is reasonably reliable, but the reliability diminishes as one proceeds from broad, inclusive class categories to narrower, more specific ones. As finer discriminations are called for, accuracy in diagnosis becomes increasingly difficult. Since this latter characteristic appears to be common to the classificatory efforts in many areas of knowledge, it would appear to be inappropriate to criticize psychiatric diagnosis on the grounds that it is less than perfectly reliable. This should not lead to an underestimation of the importance of reliability. While certain extraclassificatory factors, e.g., proficiency of the clinicians, biases of the particular clinical settings, etc., may influence it, reliability is primarily related to the precision with which classes of a schema are defined. Since the defining characteristic of most classes in psychiatric diagnosis is the occurrence of symptoms in particular combinations, the reliability of the system mirrors the specificity with which the various combinations of symptoms (syndromes) have been spelled out. It is mandatory for a classificatory schema to be reliable since reliability refers to the definiteness with which phenomena can be ordered to classes. If a system does not allow for such a division of phenomena, it can make no pretense of being a classificatory schema.

While reliability is a prerequisite if the diagnostic system is to have any value, it must not be assumed that if human effort were to make the present system perfectly reliable, it could escape all the difficulties attributed to it. This perfect reliability would only mean that individuals within each class shared a particular commonality in relation to the classificatory principle of symptom manifestation. If one were interested in attributes unrelated or minimally related to

the classificatory principle employed, the perfect reliability of the system would offer little cause for rejoicing. Perfect reliability of the present system can only be the goal of those who are interested in nothing more than the present classificatory principle and the particular attributes of the classes constructed on the basis of this principle.

When attention is shifted from characteristics which define a class to the correlates of class membership, this implies a shift in concern from the reliability of a system to its validity. The distinction between the reliability and validity of a classificatory system would appear to involve certain conceptual difficulties. It is perhaps this conceptual difficulty which explains why the rather imposing body of literature concerned with diagnosis has been virtually silent on the question of the validity of the present system of psychiatric diagnosis. Only one group of investigators (Hunt, 1951; Hunt, Wittson, & Barton, 1950a, 1950b; Hunt, Wittson, & Hunt, 1953; Wittson & Hunt, 1951) has specifically been concerned with the predictive efficacy of diagnoses and, thus, to the validity of psychiatric classifications; and even in this work, the distinction between validity and reliability is not clearly drawn.

In order to grasp the distinction between the reliability and the validity of a classificatory schema, one must differentiate the defining characteristics of the classes from the correlates of the classes. In the former case, we are interested in the principles upon which classes are formed; in the latter, in the predictions or valid statements that can be made about phenomena once they are classified. The difficulty lies in the overlap between the classifying principles and the class correlates. If a classificatory system is reliable, it is also valid to the extent that we can predict that the individuals within a class will exhibit certain characteristics, namely, those behaviors or attributes which serve to define the class.

It is the rare class, however, that does not connote correlates beyond its defining characteristics. The predictions associated with class membership may vary from simple extensions of the classificatory principles to correlates which would appear to have little connection with these principles. Let us examine a simple illustration and see what follows from categorizing an individual. Once an individual has been classified as manifesting a manic-depressive reaction, depressed type, on the basis of the symptoms of depression of mood, motor retardation, and stupor (American Psychiatric Association, 1952), the prediction may be made that the individual will spend a great deal of time in bed, which represents an obvious extension of the symptom pattern. One may also hypothesize that the patient will show improvement if electroshock therapy is employed. This is a correlate which has little direct connection with the symptoms themselves. These predictions are open to test, and evidence may or may not be found to support them. Thus, measures of validity may be obtained which are independent of the reliability of the system of classification.

The problem of validity lies at the heart of the confusion which surrounds psychiatric diagnosis. When the present diagnostic schema is assailed, the common complaint is that class membership conveys little information beyond the gross symptomatology of the patient and contributes little to the solution of the pressing problems of etiology, treatment procedures, prognosis, etc. The criticism that class membership does not predict these important aspects of a disorder appears to be a legitimate one. This does not mean the present system has no validity. It simply indicates that the system may be valid in respect to certain correlates but invalid in respect to others. Much confusion would be dispelled if as much care were taken in noting the existing correlates of classes as is taken in noting the classificatory principles. A great deal of effort has gone into the formalization of the defining characteristics of classes (American Psychiatric Association, 1952), but one looks in vain for a formal delineation of the extraclassificatory attributes and correlates of class membership. As a result, the various diagnostic categories have been burdened with correlates not systematically derived from a classificatory principle but which were attributed to the classes because they were the focal points of clinical interest. A major question is just what correlates can justifiably be attributed to the class categories. To answer this question we must turn our

attention to the purposes and philosophy underlying contemporary psychiatric diagnosis.

Philosophy and Purpose of Conventional Diagnosis

The validity of the conventional diagnostic system is least ambiguous and most free from potential criticism as a descriptive schema, a taxonomy of mental disorders analogous to the work of Ray and Linnaeus in biology. In this sense, class membership confirms that the inclusion of an individual within a class guarantees only that he exhibit the defining characteristics of that class. Only a modest extension of this system, in terms of a very limited number of well established correlates, makes for a system of impressive heuristic value, even though it falls considerably short of what would now be considered an optimal classificatory schema. As has been noted (Caveny et al., 1955; Hunt et al., 1953), the present diagnostic system is quite useful when evaluated in terms of its administrative and, to a lesser extent, its preventive implications. Caveny et al. (1955) and Wittenborn, Holzberg, and Simon (1953) should be consulted for a comprehensive list of such uses, but examples would include legal determination of insanity, declaration of incompetence, type of ward required for custodial care, census figures and statistical data upon which considerable planning is based, screening devices for the military services or other agencies, etc. In view of the extensive criticism of contemporary diagnosis, the surprising fact is not that so few valid predictions can be derived from class membership, but that so many can.

The value of the present psychiatric classification system would be further enhanced by its explicit divorcement from its Kraepelinian heritage by an emphasis on its descriptive aspect and, through careful empirical investigation, the cataloging of the reliable correlates of its categories. That this catalog of correlates would be an impressive one is expressed in Hoch's (1953) view that the present system is superior to any system which has been evolved to replace it. It is an open question whether the system merits this amount of praise. In general, however, the defense of the present system—or, for that matter, diagnosis in general (Caveny et al., 1955; Eysenck, 1952; Hunt et al., 1953; Jellinek, 1939)—tends to rest on the merits of its descriptive, empirical, and nondynamic aspects.

The present classificatory system, even as a purely descriptive device, is still open to a certain degree of criticism. Its classificatory principle is organized primarily about symptom manifestation. This would be adequate for a descriptive system if this principle were consistently applied to all classes of the schema and if the symptoms associated with each diagnostic category were clearly specified. There is some question, however, whether the system meets these requirements (Phillips & Rabinovitch, 1958; Rotter, 1954). The criticism has been advanced that the present system is based on a number of diverse principles of classification. Most classes are indeed defined by symptom manifestation, but the organic disorders, for example, tend to be identified by etiology, while such other factors as prognosis, social conformity, etc. are also employed as classificatory principles. This does not appear, however, to be an insurmountable problem, for the system could be made a completely consistent one by explicitly defining each category by the symptoms encompassed. The system would appear to be eminently amenable to the unitary application of this descriptive classificatory principle, for there are actually few cases where classes are not so defined. Where reliable relations between the present categories and etiology and prognosis have been established, these also could be incorporated explicitly within the system. Etiology and prognosis would be treated not as inherent attributes of the various classifications, but rather as correlates of the particular classes to which their relationship is known. They would, thus, not be confounded with the classificatory principle of the system.

This course of action would satisfy the requirement of consistency in the application of the classificatory principle. A remaining area of ambiguity would be the lack of agreement in what constitutes a symptom. In physical medicine, a clear distinction has been made between a symptom, which is defined as a subjectively experienced abnormality, and a sign, which is considered an objective indication of abnormality (Holmes, 1946). This differentiation has not, however, been extended

to the sphere of mental disorders. A source of difficulty may lie in the definition of what is psychologically abnormal. In psychiatric terminology, symptoms include a wide range of phenomena from the grossest type of behavior deviation, through the complaints of the patient, to events almost completely inferential in nature. One suggestion (Yates, 1958) has been to eliminate the term "symptom" and direct attention to the manifest responses of the individual. This suggestion appears to be embodied in the work of Wittenborn and his colleagues (Wittenborn, 1951, 1952; Wittenborn & Bailey, 1952; Wittenborn & Holzberg, 1951; Wittenborn et al., 1953; Wittenborn & Weiss, 1952). Wittenborn's diagnostic system, in which symptoms are defined as currently discernible behaviors, represents a standard of clarity for purely descriptive systems of psychiatric classification. This clarity was achieved by clearly noting and limiting the group of behaviors which would be employed in the system. But even here a certain amount of ambiguity remains. The number of responses or discernible behaviors which may be considered for inclusion within a diagnostic schema borders on the infinite. The question arises, then, as to how one goes about the selection of those behaviors to be incorporated in the classificatory system. Parsimony demands that only "meaningful" items of behavior be chosen for inclusion, and this selective principle has certainly been at work in the construction of all systems of diagnosis. In this sense, the present method of psychiatric classification is not a purely descriptive one, nor can any classification schema truly meet this criterion of purity. Meaning and utility inevitably appear among the determinants of classificatory systems.

Several investigators (Cameron, 1953; Jellinek, 1939; Magaret, 1952) have stressed the inappropriateness of discussing diagnosis in the abstract, pointing out that such a discussion should center around the question of "diagnosis for what?" Indeed, a diagnostic system cannot be described as "true" or "false," but only as being useful or not useful in attaining prescribed goals. Therefore, when a system is devised, its purposes should be explicitly stated so that the system can be evaluated in terms of its success or failure in attaining these objectives. Furthermore, these goals should be kept explicit throughout the period during which the system is being employed. The present diagnostic schema has not met this requirement. Instead, its goals have been carried along in an implicit manner and have been allowed to become vague. The result has been that some see the purpose of the schema as being an adequate description of mental disorders (Hunt et al., 1953), others view it as being concerned with prognosis (Hoch, 1953), and still others view the schemata goal as the discovery of etiology (Cameron, 1953).

Typically, the present schema has been conceptualized as descriptive in nature, but a brief glance at its history indicates that the original purposes and goals in the construction of this schema went far beyond the desire for a descriptive taxonomy. As Zilboorg and Henry (1941) clearly note, Kraepelin not only studied the individual while hospitalized, but also the patient's premorbid history and post-hospital course. His hope was to make our understanding of all mental disorders as precise as our knowledge of the course of general paresis. He insisted on the classification of mental disorders according to regularities in symptoms and course of illness, believing this would lead to a clearer discrimination among the different disease entities. He hoped for the subsequent discovery of a specific somatic malfunction responsible for each disease. For Kraepelin, then, classification was related to etiology, treatment, and prognosis. Had the system worked as envisaged, these variables would have become the extra-classificatory attributes of the schema. When matched against this aspiration, the present system must be considered a failure since the common complaint against it is that a diagnostic label tells us very little about etiology, treatment, or prognosis (Miles, 1953). However, it would be erroneous to conclude that the present system is valueless because its classes are only minimally related to etiology and prognosis.

What should be noted is that etiology and prognosis, though important, are but two of a multitude of variables of interest. The importance of these variables should not obscure the fact that their relationship to a classifica-

tory system is exactly the same as that of any other variables. This relationship may take one of two forms. Etiology and prognosis may be the correlates of the classes of a diagnostic system which employs an independent classificatory principle like symptom manifestation. Optimally, we should prefer a classificatory schema in which the indices of etiology and preferred modes of treatment would be incorporated (Hunt et al., 1953; Pepinsky, 1948). In essence, this was Kraepelin's approach, and it continues to underlie some promising work in the area of psychopathology. Although Kraepelin's disease concept is in disrepute (Hoch & Zubin, 1953; Marzoff, 1947; Rotter, 1954), it is the opinion of several investigators (Eysenck, 1953; Phillips & Rabinovitch, 1958; Wittenborn et al., 1953) that further work employing the descriptive symptomatic approach could well lead to a greater understanding of the etiology underlying abnormal "processes."

Another manner in which etiology, treatment, or prognosis could be related to a classificatory schema is by utilizing each of these variables as the classificatory principle for a new diagnostic system. For instance, we might organize patients into groups which respond differentially to particular forms of treatment like electroshock, drugs, psychotherapy, etc. The new schemata which might be proposed could be of considerable value in respect to certain goals but useless in regard to others. Since we do not possess a diagnostic system based on all the variables of clinical interest, we might have to be satisfied with the construction of a variety of diagnostic systems, each based on a different principle of classification. These classificatory techniques would exist side by side, their use being determined by the specific objectives of the diagnostician.

Etiology versus Description in Diagnosis

The classical Kraepelinian classification schema shows two major characteristics: a commitment to a detailed description of the manifest symptomatic behaviors of the individual and an underlying assumption that such a descriptive classification would be transitory, eventually leading to and being replaced by a system whose classificatory principle was the etiology of the various mental disorders. Major criticism of this classificatory effort has been directed at the first of these. The reservations are that, in practice, such a descriptive effort allows no place for a process interpretation of psychopathology and that it has not encouraged the development of prevention and treatment programs in the mental disorders.

The authors do not feel that the failure of the Kraepelinian system has demonstrated the futility of employing symptoms as the basis for classification. It does suggest that if one approaches the problem of description with an assumption as to the necessary correlates of such descriptions, then the diagnostic system may well be in error. Kraepelin's empiricism is contaminated in just this way. For example, he refused to accept as cases of dementia praecox those individuals who recovered from the disorder, since he assumed irreversibility as a necessary concomitant of its hypothesized neurophysiological base. Bleuler, on the other hand, who was much less commited to any particular form of causality in this illness, readily recognized the possibility of its favorable outcome. It is not, then, the descriptive approach itself which is open to criticism, but description contaminated by preconception. An unfettered description of those schizophrenics with good prognosis in contrast to those with poor prognosis reveals clear differences in the symptom configuration between these kinds of patients (Farina & Webb, 1956; Phillips, 1953).

Kraepelin's basic concern with the problem of etiology has remained a focus of efforts in the clinical area. Although his postulate of central nervous system disease as the basis of mental disorder is in disrepute, and his systematic classificatory efforts are assailed, one nevertheless finds a striking congruence between Kraepelin's preconceptions and certain current attempts at the solution of the problem of psychopathology. There is an unwavering belief that some simple categorical system will quickly solve the mysteries of etiology. The exponents of these newer classificatory schemata have merely replaced symptoms by other phenomena like test scores (King, 1954), particular patterns of interpersonal relations (Leary & Coffey,

1955), etc. It is the authors' conviction that these new efforts to find short-cut solutions to the question of etiology will similarly remain unsuccessful. The amount of descriptive effort required before etiological factors are likely to be discovered has been underestimated (Kety, 1959a, 1959b), and the pursuit of etiology should represent an end point rather than a beginning for classificatory systems. The process of moving from an empirical orientation to an etiological one is, of necessity, inferential and therefore susceptible to the myriad dangers of premature inference. We propose that the greatest safeguard against such prematurity is not to be found in the scrapping of an empirical descriptive approach, but in an accelerated program of empirical research. What is needed at this time is a systematic, empirical attack on the problem of mental disorders. Inherent in this program is the employment of symptoms, broadly defined as meaningful and discernible behaviors, as the basis of a classificatory system. Rather than an abstract search for etiologies, it would appear more currently fruitful to investigate such empirical correlates of symptomatology as reactions to specific forms of treatment, outcome in the disorders, case history phenomena, etc.

The pervasive concern with etiology may derive from a belief that if this were known, prevention would shortly be forthcoming, thus, making the present complex problems of treatment and prognosis inconsequential. Unfortunately, efforts to short-circuit the drudgery involved in establishing an empirically founded psychiatry has not resulted in any major breakthroughs. Etiology is typically the last characteristic of a disorder to be discovered. Consequently, we would suggest the search for etiology be put aside and attempted only when a greater number of the correlates of symptomatic behaviors have been established.

The authors are impressed by the amount of energy that has been expended in both attacking and defending various contemporary systems of classification. We believe that a classificatory system should include any behavior or phenomenon that appears promising in terms of its significant correlates. At this stage of our investigations, the system employed should be an open and expanding one, not one which is closed and defended on conceptual grounds. Systems of classification must be treated as tools for further discovery, not as bases for polemic disputation.

As stated above, it is possible that a number of systems of classification may be needed to encompass the behaviors presently of clinical interest. It may appear that the espousal of this position, in conjunction with a plea for empirical exploration of the correlates of these behaviors, runs headlong into a desire for conceptual neatness and parsimony. It may be feared that the use of a number of classificatory systems concurrently, each with its own correlates, may lead to the creation of a gigantic actuarial table of unrelated elements. However, the authors do not feel that such a fear is well founded because it assumes that the correlates of these systems have no eventual relation one to the other.

We believe that this latter view is unnecessarily pessimistic. While in principle a multiplicity of classificatory systems might be called for, results from the authors' own research program suggests that a single, relatively restricted and coherent classification system can be derived from an empirical study of the correlates of symptomatic behaviors (Phillips & Rabinovitch, 1958; Zigler & Phillips, 1960). Such a system might serve a number of psychiatrically significant functions, including the optimum selection of patients for specific treatment programs and the prediction of treatment outcomes. In conclusion, a descriptive classificatory system appears far from dead, and if properly employed, it can lead to a fuller as well as a more conceptually based understanding of the psychopathologies.

REFERENCES

ALLPORT, G. *Personality: A psychological interpretation.* New York: Holt, 1937.

ALLPORT, G. Personalistic psychology as science: A reply. *Psychol. Rev.*, 1946, **53**, 132–135.

AMERICAN PSYCHIATRIC ASSOCIATION, Mental Hospital Service, Committee on Nomenclature and Statistics of the American Psychiatric Association. *Diagnostic and statistical manual: Mental disorders.* Washington, D. C.: APA, 1952.

ASH, P. The reliability of psychiatric diagnosis. *J. abnorm. soc. Psychol.*, 1949, **44**, 272–277.

BEACH, F. The snark was a boojum. *Amer. Psychologist*, 1950, **5**, 115–124.

BECK, S. The science of personality: Nomothetic or idiographic? *Psychol. Rev.*, 1953, **60**, 353–359.

BOISEN, A. Types of dementia praecox: A study in psychiatric classification. *Psychiatry*, 1938, **1**, 233–236.

BROWER, D. The problem of quantification in psychological science. *Psychol. Rev.*, 1949, **56**, 325–333.

CAMERON, D. A theory of diagnosis. In P. Hoch & J. Zubin (Eds.), *Current problems in psychiatric diagnosis*. New York: Grune & Stratton, 1953. Pp. 33–45.

CATTELL, R. *Personality and motivation structure and measurement*. New York: World Book, 1957.

CAVENY, E., WITTSON, C., HUNT, W., & HERMAN, R. Psychiatric diagnosis, its nature and function. *J. nerv. ment. Dis.*, 1955, **121**, 367–380.

CRONBACH, L. The two disciplines of scientific psychology. *Amer. Psychologist*, 1957, **12**, 671–684.

EYSENCK, H. *The scientific study of personality*. London: Routledge & Kegan Paul, 1952.

EYSENCK, H. The logical basis of factor analysis. *Amer. Psychologist*, 1953, **8**, 105–113.

EYSENCK, H. The science of personality: Nomothetic. *Psychol. Rev.*, 1954, **61**, 339–341.

FALK, J. Issues distinguishing idiographic from nomothetic approaches to personality theory. *Psychol. Rev.*, 1956, **63**, 53–62.

FARINA, A., & WEBB, W. Premorbid adjustment and subsequent discharge. *J. nerv. ment. Dis.*, 1956, **124**, 612–613.

FOULDS, G. The reliability of psychiatric, and the validity of psychological diagnosis. *J. ment. Sci.*, 1955, **101**, 851–862.

GUTHRIE, E. The status of systematic psychology. *Amer. Psychologist*, 1950, **5**, 97–101.

HARLOW, H. Mice, monkeys, men, and motives. *Psychol. Rev.*, 1953, **60**, 23–32.

HARROWER, MOLLY. (Ed.) *Diagnostic psychological testing*. Springfield, Ill.: Charles C Thomas, 1950.

HOCH, P. Discussion. In P. Hoch & J. Zubin (Eds.), *Current problems in psychiatric diagnosis*. New York: Grune & Stratton, 1953. Pp. 46–50.

HOCH, P., & ZUBIN, J. (Eds.) *Current problems in psychiatric diagnosis*. New York: Grune & Stratton, 1953.

HOLMES, G. *Introduction to clinical neurology*. Edinburgh: Livingstone, 1946.

HUNT, W. Clinical psychology—science or superstition. *Amer. Psychologist*, 1951, **6**, 683–687. (a)

HUNT, W. An investigation of naval neuropsychiatric screening procedures. In H. Gruetskaw (Ed.), *Groups, leadership, and men*. Pittsburgh, Pa.: Carnegie Press, 1951. Pp. 245–256. (b)

HUNT, W., WITTSON, C., & BARTON, H. A further validation of naval neuropsychiatric screening. *J. consult. Psychol.*, 1950, **14**, 485–488. (a)

HUNT, W., WITTSON, C., & BARTON, H. A validation study of naval neuropsychiatric screening. *J. consult. Psychol.*, 1950, **14**, 35–39. (b)

HUNT, W., WITTSON, C., & HUNT, E. A theoretical and practical analysis of the diagnostic process. In P. Hoch & J. Zubin (Eds.), *Current problems in psychiatric diagnosis*. New York: Grune & Stratton, 1953. Pp. 53–65.

JELLINEK, E. Some principles of psychiatric classification. *Psychiatry*, 1939, **2**, 161–165.

KANTOR, R., WALLNER, J., & WINDER, C. Process and reactive schizophrenia. *J. consult. Psychol.*, 1953, **17**, 157–162.

KETY, S. Biochemical theories of schizophrenia. Part I. *Science*, 1959, **129**, 1528–1532. (a)

KETY, S. Biochemical theories of schizophrenia. Part II. *Science*, 1959, **129**, 1590–1956. (b)

KING, G. Research with neuropsychiatric samples. *J. Psychol.*, 1954, **38**, 383–387.

KOCH, S. The current status of motivational psychology. *Psychol. Rev.*, 1951, **58**, 147–154.

LEARY, T., & COFFEY, H. Interpersonal diagnosis: Some problems of methodology and validation. *J. abnorm. soc. Psychol.*, 1955, **50**, 110–126.

MACKINNON, D. Fact and fancy in personality research. *Amer. Psychologist*, 1953, **8**, 138–146.

MAGARET, ANN. Clinical methods: Psychodiagnostics. *Annu. Rev. Psychol.*, 1952, **3**, 283–320.

MARQUIS, D. Research planning at the frontiers of science. *Amer. Psychologist*, 1948, **3**, 430–438.

MARZOFF, S. S. The disease concept in psychology. *Psychol. Rev.*, 1947, **54**, 211–221.

MASLOW, A. Cognition of the particular and of the generic. *Psychol. Rev.*, 1948, **55**, 22–40.

MEHLMAN, B. The reliability of psychiatric diagnosis. *J. abnorm. soc. Psychol.*, 1952, **47**, 577–578.

MENNINGER, K. The practice of psychiatry. *Dig. Neurol. Psychiat.*, 1955, **23**, 101.

MILES, H. Discussion. In P. Hoch & J. Zubin (Eds.), *Current problems in psychiatric diagnosis*. New York: Grune & Stratton, 1953. Pp. 107–111.

NOYES, A. *Modern clinical psychiatry*. Philadelphia: Saunders, 1953.

PATTERSON, C. Is psychotherapy dependent on diagnosis? *Amer. Psychologist*, 1948, **3**, 155–159.

PEPINSKY, H. B. Diagnostic categories in clinical counseling. *Appl. psychol. Monogr.*, 1948, No. 15.

PHILLIPS, L. Case history data and prognosis in schizophrenia. *J. nerv. ment. Dis.*, 1953, **117**, 515–525.

PHILLIPS, L., & RABINOVITCH, M. Social role and patterns of symptomatic behaviors. *J. abnorm. soc. Psychol.*, 1958, **57**, 181–186.

RAPAPORT, D. The future of research in clinical psychology and psychiatry. *Amer. Psychologist*, 1947, **2**, 167–172.

ROBY, T. An opinion on the construction of behavior theory. *Amer. Psychologist*, 1959, **14**, 129–134.

ROE, ANNE. Integration of personality theory and clinical practice. *J. abnorm. soc. Psychol.*, 1949, **44**, 36–41.

ROGERS, C. Significant aspects of client-centered therapy. *Amer. Psychologist*, 1946, **1**, 415–422.

ROGERS, C. *Client-centered therapy*. Boston: Houghton Mifflin, 1951.

ROTTER, J. *Social learning and clinical psychology*. New York: Prentice-Hall, 1954.

SCHMIDT, H., & FONDA, C. The reliability of psychiatric diagnosis: A new look. *J. abnorm. soc. Psychol.*, 1956, **52**, 262–267.

SCOTT, J. The place of observation in biological and psychological science. *Amer. Psychologist*, 1955, **10**, 61–63.

SCOTT, W. Research definitions of mental health and mental illness. *Psychol. Bull.*, 1958, **55**, 1–45.

SEEMAN, W. Psychiatric diagnosis: An investigation of interperson-reliability after didactic instruction. *J. nerv. ment. Dis.*, 1953, **118**, 541–544.

SKAGGS, E. Personalistic psychology as science. *Psychol. Rev.*, 1945, **52**, 234–238.

SKAGGS, E. Ten basic postulates of personalistic psychology. *Psychol. Rev.*, 1947, **54**, 255–262.

THORNE, F. Back to fundamentals. *J. clin. Psychol.*, 1953, **9**, 89–91.

TOLMAN, R. Virtue rewarded and vice punished. *Amer. Psychologist*, 1953, **8**, 721–733.

TYLER, LEONA. Toward a workable psychology of individuality. *Amer. Psychologist*, 1959, **14**, 75–81.

WINDLE, C., & HAMWI, V. An exploratory study of the prognostic value of the complex reaction time tests in early and chronic psychotics. *J. clin. Psychol.*, 1953, **9**, 156–161.

WITTENBORN, J. Symptom patterns in a group of mental hospital patients. *J. consult. Psychol.*, 1951, **15**, 290–302.

WITTENBORN, J. The behavioral symptoms for certain organic psychoses. *J. consult. Psychol.*, 1952, **16**, 104–106.

WITTENBORN, J., & BAILEY, C. The symptoms of involutional psychosis. *J. consult. Psychol.*, 1952, **16**, 13–17.

WITTENBORN, J., & HOLZBERG, J. The generality of psychiatric syndromes. *J. consult. Psychol.*, 1951, **15**, 372–380.

WITTENBORN, J., HOLZBERG, J., & SIMON, B. Symptom correlates for descriptive diagnosis. *Genet. psychol. Monogr.*, 1953, **47**, 237–301.

WITTENBORN, J., & WEISS, W. Patients diagnosed manic-depressive psychosismanic state. *J. consult. Psychol.*, 1952, **16**, 193–198.

WITTMAN, P., & SHELDON, W. A proposed classification of psychotic behavior reactions. *Amer. J. Psychiat.*, 1948, **105**, 124–128.

WITTSON, C., & HUNT, W. The predictive value of the brief psychiatric interview. *Amer. J. Psychiat.*, 1951, **107**, 582–585.

YATES, A. Symptoms and symptom substitution. *Psychol. Rev.*, 1958, **65**, 371–374.

ZIGLER, E., & PHILLIPS, L. Social effectiveness and symptomatic behaviors. *J. abnorm. soc. Psychol.*, 1960, **61**, 231–238.

ZILBOORG, G., & HENRY, G. W. *History of medical psychology*. New York: Norton, 1941.

(Received December 22, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 63, No. 3, 619-628

MUTUAL ACHIEVEMENT STRIVINGS:

A CONTINUUM FOR MENTAL HEALTH

PHILIP LICHTENBERG
Graduate Department of Social Work and Social Research, Bryn Mawr College

RHONDDA K. CASSETTA
New York State Mental Health Research Unit

AND JOHN C. SCANLON
System Development Corporation

THIS report pertains to a description of mental health and disorder that emphasizes a continuum called Mutual Achievement Health. The analytic system is derived partly from past work (Lichtenberg, 1955; Lichtenberg, Kohrman, & MacGregor, 1960), partly from attempts to classify recent events in the lives of newly admitted hospital patients, and partly from efforts to coordinate a mental health theory with a formulation of types of treatment (Lichtenberg, Cassetta, & Scanlon, 1960a).

It is our belief, as well as that of recent reviewers (Herzog, 1959; Jahoda, 1958; Scott, 1958a, 1958b), that a good theory of mental health is vital for research and practice in the fields of mental health education, the epidemiology of mental disorder, and the evaluation of psychotherapy or other treatment services. It is also our belief, as it is Maslow's (1954), that the study of psychologically disordered persons in the therapeutic setting is strategic for the development of scientific ideas about what are good human relationships and what is the good society. This report is intended to contribute to these fields and endeavors.

GENERAL THEORETICAL CONSIDERATIONS

Apart from questions about the value of conceptualizing mental health (e.g., Freides, 1960; Szasz, 1960), three views regarding mental health and disorder are vying for scientific dominance at present:

1. Persons can be measured on a continuum according to their distance from complete mental health or complete mental disorder.

2. Two dimensions exist—one defining the extent of mental health, the other defining the extent of mental illness. For example, Korner (1958) proposed two continua, "mental health-mental injury" and "susceptibility to mental illness-mental illness."

3. There is no quantitative factor associated with mental disorder; all differences are qualitative.

We prefer the continuum thesis because it seems more consistent with fundamental scientific procedure and because the alternatives seem inadequate. To divide matters into two continua (positive mental health and mental illness) seems to conceal the judgment of severity on a single continuum. In practice, one must decide which of the two dimensions is more appropriate for a given individual, a choice that represents acting according to a judgment of severity. If one evaluates all people on both dimensions, the problems of overlap negate the value of separate dimensions. To propose that one study only qualitative differences, such as diagnostic classes, similarly postpones unduly a severity measurement. We inevitably act as if there are severity differences when we plan treatments for schizophrenics or anxiety neurotics. These alternatives will be subsumed within a single mental health dimension in the long run in any case.

Two assumptions other than the continuum thesis have guided this work. First, we have supposed that conscious experience, especially in the form of intentions, is *companion* to the social actions of the person and must be so conceived. We are proposing to observe simultaneously social actions and private intentions, noting their congruencies and discrepancies. We treat intentions as parts of actions, not motivators of actions. Thus, we assume that one must know what a

person intends to do, what he has in mind, and that one must also know how he correlates his intentions with his actions. The investigation of many intentions-in-transactions incorporates the necessary and sufficient conditions for evaluating mental health.[1]

The second assumption is that environmental contributions to transactions must be evaluated when estimation is made of mental health level. It is one of the more significant facets in our theory that environmental contributions are considered to vary and to do so systematically. For one thing, environmental agents may be more or less inducing (initiating) or responsive (reacting) with respect to the person being studied. For another, environmental agents may encourage inadequate intentions and discourage superior ones (or vice versa). Both of these influences act to prevent any person from rising far above his fellow men with respect to mental health level. Conversely, the most disturbed among us cannot be so very different from the rest of us, for his disturbance is partly reflective of our actions as environmental forces.

OUR MENTAL HEALTH THEORY

Mental health is a quality of personality, reflected in virtually every action of a person's life. We mean by this, first, that mental health is an attribute of a personality system. Second, by the time a person possesses a definite personality system (sometime during childhood or early adolescence), he has also a discernible mental health level that is relatively persistent. Third, the mental health factor affects all actions of a person and is, thus, open to observation at all times. It takes exceedingly microscopic observation, however, to detect the entire structure in one or two events. It is more reliable to observe consistent biases and their central tendency or adaptation point over a variety and number of events. Momentary influences are muted in favor of the consistencies of the personality system. Thus, mental illness is considered to be a generalized tendency of a person to deviate to a given degree from what is "appropriate" (a matter to be discussed later). Further, mental health pertains to psychological events in which the person is actively helping to create the event, to give it content, scope and meaning. The events are transactional.

While the person is never entirely responsible for the totality of an event, he is responsible for his intentions. We take intentions to mean the purposes, plans, or projected outcomes that the person consciously experiences, that he has in mind. An intention is what a person "wants" to do when he knows that he "wants" to do it. A person who turns on the gas in the oven and puts his head in may want to die, or he may want to be rescued. One must study his phenomenological experience to discover which is his intention.

The aspect of the intentions that seems significant to a theory of mental health is whether the intention reflects strivings toward equality or inequality. The equality factor refers to the sharing of whatever satisfactions accrue to the transactions. Equality is intended if the person and the others are supposed to suffer or gain together and in roughly equal proportions. Inequality takes either an aggrandizing or a sacrificing form. In the aggrandizing intention, the imbalance favors the person; in the sacrificing intention, the imbalance favors the other.

Because intention and action need not coincide, evaluating intentions solely is insufficient for determining mental health. When a person is intending equality, he may also be creating circumstances that inhibit the practical realization of equality. He may

[1] The influence of the unconscious is inferred when there is a discrepancy between intention and transaction and when that discrepancy is not determined primarily by unforeseen contributions from the environment. This is to say, before any inference can be drawn about unconscious forces, one must first note a discrepancy and then note that it was primarily a function of the person.

This analysis suggests two types of situations in which interpretations of unconscious motivation are irrelevant: Situations in which the intentions and transactions are congenial so that the intention is fulfilled. The overall intention-transaction symmetry (given even minor discrepancies) makes reference to unconscious motivation trivial. Such agreement of intention and transaction is probably more frequent than we have been led to believe by the emphasis on man's irrationality. Situations in which a person's environment is primarily responsible for the discrepancy between intention and transaction.

reveal a discrepancy between his theory and his practice, his thought and his action.

The feature of events that seems significant in providing a context for the intention toward equality or inequality is a transactional factor here called *mutual achievement*. All events are characterized, on the one hand, by some goal or purpose which gives expression to the goal directed or *achievement* character of the event. (We speak here of the goals of the transaction, not the intention of the person.) All events contain, on the other hand, some indication of the sharing among the participants of the achievements gained, of the implications of their mutual inclusiveness or exclusiveness, in brief, of the *mutuality* of satisfaction.

We contend that the achievement factor and the mutuality factor are always integrated in events and that one observes only mutual achievement strivings. All behaviors must be observed, accordingly, with simultaneous attention to achievement and sharing factors. For instance, a person's orientation toward achievement is determined in part by his subjective probability of success and in part by his anticipation of how that which is achieved is to be divided among the collaborators. Similarly, a person's sharing propensities will depend both on his anticipations of mutuality and his expectancy of successful goal attainment (cf. Lichtenberg, Kohrman, & MacGregor, 1960, pp. 200–203).

We have formulated four distinguishable levels of mutual achievement strivings. At the lowest level are to be found minimal mutuality and achievement. Whether in examining persons on this level one emphasizes the mutuality component (e.g., a man as a social isolate) or the achievement component (e.g., a man as a poor worker), one sees that the person's attempts at relating both to people and to goals lead to loneliness and to low level accomplishments. At the highest level, of course, the reverse is true. In describing these four levels, we proceed from the highest to the lowest, from that presumed more healthy to that presumed less healthy.

The highest level of mutual achievement striving in this system (Level IV) is called the "Redefine" level. In events at this level, the person's intentions are met with attempts at redefining or restructuring by the other(s) in the interaction. The intentions of the person are affirmed as starting points, incomplete and awaiting development, while at the same time revision and expansion of these intentions or their structural arrangements is already taking place. The other in the transaction is suggesting by his manner or actions that the person's intentions are *tentatively* acceptable, but acceptable primarily as origins from which will grow new intentions, ones that will permit closer integration of the person and the other. The other in the transaction aims to restructure the person's intentional experience so that his own intentions and wishes are given more expression. The alignment of the person's intentions vis-à-vis the actions of the other is presented in Figure 1.

The redefining process may focus upon the goal, the proposed sharing of rewards and punishments, the coordination of efforts, or other similar manifestations of mutual achievement striving. The focus will vary according to the actual congruence between the person and the other in their intentions, their past experiences in coordinated effort, the nature of the task, and other such "reality" factors. When there is much willingness to share rewards, more of the redefining endeavor will center upon the goals to be sought or the ways in which goals may be reached. When there are

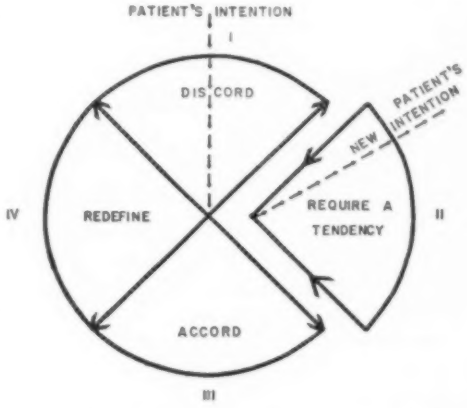

FIG. 1. Representation of maneuvers by the other in a transaction vis-à-vis a person's intentions with reference to levels of mutual achievement strivings.

common goals, the restructuring concerns sharing, cooperation, and coordination.

The next level of mutual achievement striving (Level III) is that called the "Accord" level. In events labeled "Accord," the person's intentions are affirmed by the other, not only tentatively as starting points, but decisively, as the intentions whose development will control the future. The goals of the person in their original form become the goals of the transaction. The initial sharing patterns are accepted without major attempts at modification. The other in the transaction is receptive to or promotes selected intentions of the person (those that are congenial, of course), to which he then adds his talents and skills and hopes. The other aligns himself in accord with the person's direction, as indicated in Figure 1. It represents that form of cooperation in which I do what you want, thereby cooperating with you.

At this level there is typically an emphasis on maximum mutuality between the person and the other; but this is not the same as maximum mutual achievement striving. The mutuality alone is divorced to a certain extent from the goals that the other may wish to posit. The other in the transaction is forced to inject his intentions in an indirect way—by *selectively* attending to the intentions offered to him. He aligns himself only with those intentions which promise gratification to him as well as to the originator. Some constructive achievement components are, thus, lost to the transaction. They are transformed from task oriented achievement striving to endeavors for the attainment of mutuality as achievement itself.

Level II in mutual achievement strivings contains those events in which the person meets a call for new intentions, that is, for intentions not currently active or likely to be active were he alone to dictate the course of events. The other in these transactions tries to induce distinctly new intentions into the person's life. It is asking the person to "want" in a way he has not previously conceptualized.

It is true that all actions by the other tend to call forth responses from the person and in some sense to require new behaviors and intentions. If one redefines another's behavior, for example, he accepts the original intentions only as stations on the road to something new and different, as places to obtain fuel that will be consumed as movement progresses and newly created intentions arise. We do not use the Level II mutual achievement analysis, however, to refer to these types of new intentions being required. When something new is called for directly and distinctly, it is to be recognized in that form. Sometimes the clue is expressed by surprise, perplexity, or hestitation before any commitment is made. Whatever the clue, it does not seem difficult to infer this type of transaction.

The remaining level of mutual achievement striving (Level I) includes events in which the other aligns himself in discord with the person's intentions or selects goals that will not be acceptable to the person. There is, of course, a minimum of sharing and of achievement in this level. The other in the transaction aims to inhibit the development of the expressed intentions; but because a relationship is presumed to exist for an event to be observed, efforts are being expended to create a different set of intentions in the person through the rejection of those expressed.

The other implies that he wants the relation to move to a new plane of intentions. One way, of course, is to have the person-other relation dissolved. Another way, and probably the expected way as far as the opposing other experiences it, is for the person to inject a radically different intention into the transaction, one that will be congenial so that the relationship and goal striving can progress.

These are the four levels of mutual achievement striving that we discern: Redefine (IV), Accord (III), Require a Tendency (II), and Discord with the person's intention by the other (I). The differences between the levels reflect both mutuality or sharing differences and achievement or goal attainment differences.

We have developed the thesis that the important aspect of intentions is the striving toward equality or inequality. Aggrandizing and sacrificing forms of inequality are counted as equivalents for this analysis, since they do not contribute differently to severity of

TABLE 1

A QUANTIFICATION OF INTENTIONS TOWARD EQUALITY OR INEQUALITY WITHIN LEVELS OF MUTUAL ACHIEVEMENT STRIVINGS

Weighted score	The intention-transaction combination
1	An *equalitarian* intention is *redefined*
2	An *equalitarian* intention finds *accord*
3	An *equalitarian* intention follows a *requirement of a tendency*[a]
3	An *aggrandizing* or *sacrificing* intention is *redefined*
4	An *equalitarian* intention finds *discord*
4	An *aggrandizing* or *sacrificing* intention finds *accord*
5	An *aggrandizing* or *sacrificing* intention follows a *requirement of a tendency*[a]
6	An *aggrandizing* or *sacrificing* intention finds *discord*

[a] Since "require a tendency" presumes the lack of a specific intention by the person in the determination of an event, we took the intention that developed as a function of the event for this area.

disorder. We have also proposed that the important facet of the transaction is the mutual achievement striving. Combining these separate proposals gives eight varieties of intention-transaction categories, which are presented in Table 1 in the form used to quantify each event recorded. The eight varieties include intentions toward equality within the four levels of mutual achievement striving and intentions toward inequality within the four levels.

In giving quantitative values to the eight categories, we were led to establish two equivalents. This equivalence derives from the fact that the person helps to create the transaction in addition to creating his intention. Thus, functional equivalence exists between posing an equalitarian intention in a transactional context of discord and posing an aggrandizing intention that is supported by the other. On the one hand, the discord of the transaction limits the extent of satisfaction that is to be shared equally; on the other hand, the kind of aggrandizing supported by the other is restricted by his selectivity regarding which intentions he will sustain. The aggrandizing is restricted by the conditions of support in the transaction, and the intention toward equality is limited by the lack of communion in the transaction.

Similar equivalents are an equalitarian intention that follows the requirement of a new intention by the other, and an aggrandizing or sacrificing intention that is redefined by the other. We do not credit the person with the original, private development of the intention toward equality in the first instance; we do credit him for selecting a good redefining transaction in which to express aggrandizing intentions in the second instance.

The procedure for developing an estimate of severity of mental disorder treats the mental health level as the central tendency of deviations over a variety of events. A series of events in a person's life is scored, and an average is taken. The larger the series scored, of course, the greater the reliability of such a central tendency.

Because mental disorder is conceived as a consistent deviation from something designated as "appropriate," a deviation must be a deviation from some standard. The standard, or value, or that-which-is-appropriate, in our case, is an equalitarian intention in a redefining transaction.

METHOD

This analysis of mental health was implemented in a clinical record form relevant to disposition case conferences. We collected information pertaining to significant events in the recent past of 164 patients admitted to Syracuse Psychiatric Hospital[2] between late October 1958, and early June 1959. The rating schedule, which includes items on functional disability, diagnosis, presenting complaint, type of treatment, and locomotion in treatment, was filled out by the three authors, operating in three teams of two persons each when the patients and their histories were presented by the responsible residents and interns to the director of the hospital or his assistant director. Thus, the information used is our reflection of the interviewer's understanding of the patient's story and that of his relatives, friends, welfare workers, etc. The patients were nearly always presented to the staff in person, and new material and verification or clarification from these presentations were obtained. Further information was gained from a systematic search of the written records kept by the hospital about the patients.

For a significant event to be recorded, it was necessary to have reasonable information regarding four matters: who initiated the event, the patient or the other; the environmental area—family, marital, religious, etc.—in which the other exists; the mutual achievement characteristic of the patient-other relation; and a descriptive basis for determining whether the patient's intentions were aggrandizing, sacrificing, or equalitarian.

[2] Syracuse Psychiatric Hospital is a state hospital with 60 beds, attached to the Upstate Medical Center as a teaching and research hospital and oriented toward diagnosis and short-term treatment. We are indebted to Marc Hollender, the director, for permission to work in the hospital and for the help he and his staff gave to us.

The last factor, the descriptive basis, varied according to several issues that accompanied our aim for immediate descriptiveness. We directly rated the patient's intention as toward or away from treatment in those events connected with locomotion toward the hospital. We inferred from these an equalitarian intention if the patient aimed toward, and a nonequalitarian intention if he aimed away from treatment. (Since we did not wish to overweight this sort of event, the locomotion scores were averaged, and the resulting average score was used as one of the series of scores pertaining to other events.) Regarding other events, behaviors were described briefly, and later ratings were made concerning the equalitarian-inequalitarian nature of the intentions. If the patient initiated the event, we described his behavior. If the other initiated the event, we recorded the environmental pressure and inferred the patient's intention from the equality of the pressure and the mutual achievement character of the transaction.

An "event" was taken as experienced by the patient. Using the patient's definition of an event is important for establishing a proper unit of observation. Some examples may be illuminating.

1. An ex-patient foresaw the return of his severe depressive feelings and turned to his family physician for help in returning to the hospital. The doctor arranged for his admission. We noted that (*a*) the patient initiated the transaction, (*b*) the environmental area was "other kinds of concurrent treatment," (*c*) the patient and doctor were in accord, and (*d*) the patient's intention was toward treatment. Counting the "toward treatment" as an equalitarian intention and the transaction as accord gives an intention-transaction weighted score of 2 for this event.

2. One patient consulted her priest about what she felt to be her badness about not wanting her recently born child. The priest tried to assure her (and only upset her more) by saying she had not committed a sin. We recorded that (*a*) the patient initiated the event, (*b*) the area was religion, (*c*) the patient and priest were in discord, and (*d*) the intention was to sacrifice. The denial in the transaction of a sacrificial intention was given a weighted score of 6.

3. A patient's husband, after a long period of debate, suddenly decided on his own that the family should move out of their home into a new one. He was prepared to sacrifice his own desires to please his wife. We rated this as (*a*) other initiates, (*b*) marital area, (*c*) patient-husband accord, and (*d*) patient's intention as aggrandizing (the husband's sacrifice plus accord implying the patient's intention). We scored this, accordingly, as a 4.

TABLE 2

INTERRATER AGREEMENT IN RATING INTENTIONS

Intention	Agree		Disagree[a]		Total	
	N	%	*N*	%	*N*	%
Equalitarian	69	65.7	36	34.2	105	100.0
Nonequalitarian	530	87.6	75	12.4	605	100.0
Total	599	84.4	111	15.6	710	100.0

[a] If two raters scored the intention as equalitarian and the third did not, the accounting was Equalitarian-Disagree. Similarly with two raters scoring nonequalitarian, the accounting was Nonequalitarian-Disagree. Note that this table encompasses all three raters judging each item.

TABLE 3

INTERRATER AGREEMENT IN RATING LEVEL OF MUTUAL ACHIEVEMENT

Level	Agree		Disagree		Total	
	N[a]	%	*N*	%	*N*	%
IV Redefine	16	61.5	10	38.5	26	100.0
III Accord	31	60.8	20	39.2	51	100.0
II Require a tendency	14	38.9	22	61.1	36	100.0
I Discord	122	79.2	32	20.8	154	100.0

[a] The total number of events included in this analysis is small because only events selected initially by both raters are used. This procedure enhances the "independence" of the ratings.

Reliability

The total number of scores used in estimating the mutual achievement health of the 164 patients was 874, an average of 5.33 per person. This total consists of 710 items other than locomotion and 164 locomotion average scores. The scores for each person were averaged to give a mental health score. The mental health scores ranged from 2.3 to 6.0 with a mean of 4.6 and a median of 4.7. The scores tend toward the upper limits, partly because we have a hospital population that is presumably more mentally disturbed than a nonhospital population and partly because significant events reported at a case conference tend to favor descriptions of more serious disabilities and troubles.

In examining agreement among raters, distinction is made between the selection of events from the mass of information available and the classification of selected events. Group decision prevailed in the final selection and scoring, but we wished to study how consistently different people could apply the theoretical analysis. Between 47% and 74% of events proposed by one rater were also chosen by the other; the average of the three raters taken with each other was 62%.

When events have been selected, comparison can be made of agreements on the ratings assigned to the events. The intentional and the transactional characteristics of the events were rated separately in order to minimize

such halo effects as may occur in the creation of a single score for a person from many items and in rating one sort of information along with other material with which correlations are to be made. The results are presented in Tables 2 and 3. In Table 3, the percentages of agreement correlate with an intuitive estimate of what is easiest to rate (deny or oppose) and what is hardest (requires a new tendency). The agreement in Table 2 refers to all three raters judging simultaneously; that in Table 3 is derived from a team-by-team judging process. The information in these tables lends credence to the final scores evolved in group discussion.

Tests Applied to the Theory

Two sorts of tests beyond consistency are pertinent to the scientific valuation of this theoretical analysis. First, how well do the results based on this system conform to other estimates of severity of mental disorder, and second, does placing people on this continuum relate to such practical issues as type of treatment and disposition of the patient?

For comparison purposes, we used two measures of severity of mental illness. The first of these is "Functional Disability," a theoretical scheme we developed (Lichtenberg, Cassetta, & Scanlon, 1960b) which is based on the extent of a patient's disability (none, mild, moderate, or severe) in each of four domains or foci of relations: self-others, self-things, self-body, and self-psyche. Summing the degree of disability in the four domains permits a range from a score of 0 to a score of 12. For the 164 cases reported, the observed severities range from 1.5 to 11.0 with a mean of 5.6, a median of 5.5, and a standard deviation of 1.6.

There is minimal overlap between the information determining the mutual achievement health score and that determining the functional disability score. Any correlation between the two scores is likely to be a function of their measurement of a person, not of their interdependence methodologically.

The second measure of presumed severity of mental illness which we used is diagnosis as given in the American Psychiatric Association standard nomenclature of mental disorders. The diagnostic categories applied, in order of presumed severity of illness, are schizophrenia, affective psychosis (including involutional melancholia), personality pattern disturbance, sociopathic personality disturbance, psychoneurotic reactive depression, all psychoneuroses other than reactive depression (called "other neuroses"), and transient situational personality disorder. In the comparisons that test the mutual achievement health analysis, the following are combined, either because the numbers are small or other, prior results suggest their equivalence: personality pattern disturbance plus sociopathic personality disturbance (small numbers); "other neuroses" and transient situational personality disorders (functionally equivalent), and a miscellaneous group (small numbers).

The next test to which the mutual achievement health ratings are put concerns the prescriptions of therapy. The formulation of treatment types (Lichtenberg, Cassetta, & Scanlon, 1960a) derives from the same theoretical analysis as the mutual achievement health scheme. Indeed, it is a particularization of the general theory with special reference to psychological treatment as a task and the therapist as the other in the transaction. In the functional disability study, we were able to demonstrate that severity of disorder and level of mutual achievement striving in therapy vary together. The same should hold true with the mutual achievement health ratings. Comparisons were made by the following fourfold classification, so limited because one level of treatment has few cases: no treatment or not established; uncovering and interpreting treatment; supportive care or changing the patient's world to enhance fulfillment of his wishes; and directive care or protective treatment, including the protection

TABLE 4
THE RELATION OF MUTUAL ACHIEVEMENT HEALTH SCORES AND FUNCTIONAL DISABILITY SCORES

Mutual achievement scores	Functional disability scores				
	0–4.7	4.8–5.5	5.6–6.5	6.6–11.0	Totals
0–4.2	22	12	4	6	44
4.3–4.6	12	10	18	3	43
4.7–5.1	6	14	6	10	36
5.2–6.0	3	8	10	20	41
Totals	43	44	38	39	164

$\chi^2 = 48.21$, $df = 9$, $p < .001$ C = .48

TABLE 5
THE RELATION OF MUTUAL ACHIEVEMENT HEALTH SCORES AND DIAGNOSTIC IMPRESSION

Mutual achievement scores	Diagnostic impression						
	Schizophrenia	Affective psychosis	Personality pattern and sociopathic	Neurotic depression reaction	Other neuroses[a] and situational	All other diagnoses	Total
0–4.2	5	6	2	12	14	5	44
4.3–4.6	7	8	9	7	6	6	43
4.7–5.1	14	7	5	5	2	3	36
5.2–6.0	18	6	5	2	3	7	41
Totals	44	27	21	26	25	21	164

$\chi^2 = 37.20$, $df = 15$, $p < .002$ C = .43

[a] This class includes all psychoneuroses other than depression reaction.

of the patient from himself and the protection of others.

The final test concerns what may be called outcome. While severity of disorder is not the sole determinant of a patient's disposition or ability to remain out of hospital, it is probably a significant one. Patients were divided into the following three classes: patients discharged to the community who remained out of a psychiatric hospital for at least 9 months subsequent to discharge ("out, not back"); patients discharged to the community who returned to any psychiatric hospital within 9 months after discharge ("out, back"); and patients sent directly to another inpatient psychiatric hospital for care ("to other inpatient care"). Since Syracuse Psychiatric Hospital does not ordinarily provide long-term care, these categories seem reasonable for a practical evaluation of the mental health theory. The expectation was that the first group of patients would cluster on the healthier side, the second group in the middle, and the third group on the more disordered side of a ranking of all patients on mental health.

RESULTS

The findings in the comparison between Mutual Achievement Health and Functional Disability are presented in Table 4. The distribution of patients in this table follows theoretical expectations with minor exceptions (e.g., the four central cells), where greater precision is needed to separate correctly people who are in the central cells because the range of scores defining them (e.g., 4.8–5.5 and 5.6–6.5 in Functional Disability) is narrower than the range of scores for the peripheral cells (e.g., 0–4.7 and 6.6–11.0 in Functional Disability).

The second comparison (Table 5) involves

TABLE 6
MUTUAL ACHIEVEMENT HEALTH SCORES AS RELATED TO TYPES OF TREATMENT

Mutual achievement health scores	Treatment types				
	None or not established	IV Uncover	III Support or Enhance	II or I Direct or Protect	Totals
0–4.2	7	11	15	11	44
4.3–4.6	6	4	13	20	43
4.7–5.1	5	4	7	20	36
5.2–6.0	4	3	4	30	41
Totals	22	22	39	81	164

$\chi^2 = 23.59$, $df = 9$, $p = .005$ C = .35

TABLE 7
THE RELATION OF MUTUAL ACHIEVEMENT HEALTH SCORES TO OUTCOME

Mutual achievement scores	Outcome			
	Discharged-no readmission within 9 months	Discharged-readmission within 9 months	Discharged to another hospital	Totals
0–4.2	33	6	4	43
4.3–4.6	28	11	1	40
4.7–5.1	17	11	6	34
5.2–6.0	16	9	15	40
Totals	94	37	26	157[a]

$\chi^2 = 25.47$, $df = 6$, $p < .001$ C = .37

[a] This total differs from the 164 total studied because 7 patients could not be classified in these three groups.

the Mutual Achievement Health formulation and diagnostic impression. The distribution of persons in this table also tends to confirm the theory being examined. Although there are some failures in the data, the overall tendencies are confirmatory.

The Mutual Achievement Health scores fail to give order to the group of patients diagnosed as affective psychosis. Because these patients were adequately handled in the Functional Disability investigation and because the neurotic counterpart of a psychotic affective state, neurotic depressive reaction, is in accord with theoretical expectations in the current study, this failure is perplexing.

The relation of preferred type of treatment and Mutual Achievement Health is shown in Table 6. The distribution of the 164 patients shows a statistical significance that confirms the hypothesized relation. The healthier patients tend to be given those treatments in which the level of mutual achievement strivings is higher, and the less healthy patients are given those in which the level is lower.

The final test was health level and outcome. The results are presented in Table 7. They tend to confirm the utility of the mental health theory. More of those people who are discharged and not readmitted to a hospital within 9 months are at the healthy end of the scale; those people who are discharged and readmitted cluster in the middle scores; and those who are sent directly to another hospital are more likely to be scored in the lower echelons of the Mutual Achievement Health dimension.

Inspection of the records of the 16 patients with a low mental health score who are discharged but do not return suggests that some of these people were discharged to their homes for essentially custodial care. One man died in the interim. Others were hospitalized for acute phases of chronic disorders and released when these phases were ended. In most instances, therefore, the mental health scores of these 16 people do not seem to be faulty.

DISCUSSION

A close reading of the theoretical analysis reveals that the person under observation is not classified as redefining the intentions of the other, according with them, requiring new ones, or disagreeing with the expressed intentions of the other. We do not study the person's behavior as it fits the intentions of others, which a complete examination of personality might profitably include, but only the impingement of the other on the person's intentions. That is, we do not study the person as an object in the life of the other. This limitation reflects the fact that this study was done in a clinical setting where patients' intentions are more accessible to observation than the intentions of collaborators.

A slightly different matter concerns who initiates each event, who is the "attributed source" as Heider (1958) put it. When the person's intentions are redefined, he is necessarily the initiator; when a new intention is required of him, he must be a reactor. However, he can be initiator or reactor when there is either an accord or discord level of mutual achievement striving. Thus, the four levels of mutual achievement striving expand into six types of striving.

From an examination of all the items, we conclude that six concomitant factors accompany mutual achievement levels. By considering a factor to be a concomitant, we mean both that it goes with mutual achievement strivings levels in a natural connection and that it accompanies in a subordinate capacity. These six factors are:

1. The *apparent* unification of mutuality and achievement concerns. At upper levels of mutual achievement strivings, it is difficult to separate mutuality facets from achievement facets. At lower levels, people's actions seem to be either mutuality directed or achievement directed, even though they are mutual achievement strivings in fact.

2. The number and complexity of intentions in the events. In the upper levels, many and complex intentions fit within events; in the lower levels, few and simple ones are present.

3. Reliance upon external authority for progress or maintenance of the relation. In the higher levels of mutual achievement strivings, there is little reliance upon hierarchy, tradition, or outside authority. The reverse is true at the lower levels.

4. The autonomy of the actors. At the higher

levels, the participants bring to the relation autonomously formed plans and proposals; at the lower levels, a certain emptiness is evident.

5. The combined embeddedness, openness, and sense of development characteristic of events. The events of a higher level are more directly embedded in other, often larger, events; whereas those of a lower level are discrete, being seen in larger units, mainly as repetitive acts.

6. The apparent discreteness of events as events. The determination of what sequences are "events," the units of observation, is more difficult for the higher levels than for the lower levels.

SUMMARY

This report presents a theory of mental health formed under the influence of three general considerations: persons can be described according to a position on a mental health continuum; conscious experience plus transactional characteristics of social behavior comprise the essential features for determination of mental health; and the environment is to be included in variable and systematic fashion in such a determination. From the analysis of intentions and mutual achievement strivings are derived eight quantified combinations of intentions-in-transactions. Applying the scoring system to a series of events provides a quantitative statement of mental health level.

The theory was examined with respect to the histories of 164 hospitalized patients. The outcomes were compared to an alternate analysis of mental health, to diagnosis, to treatment prescribed, and to outcome. The results tend to verify the theory.

REFERENCES

FREIDES, D. Toward the elimination of the concept of normality. *J. consult. Psychol.*, 1960, **24**, 128–133.

HEIDER, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

HERZOG, ELIZABETH. *Some guide lines for evaluative research: Assessing psycho-social change in individuals.* Washington, D. C.: United States Department Health, Education, and Welfare, Social Security Administration, Children's Bureau, 1959.

JAHODA, MARIE. *Current concepts of positive mental health.* New York: Basic Books, 1958.

KORNER, I. N. Mental health vs. mental illness. *Ment. Hyg., N. Y.*, 1958, **42**, 315–320.

LICHTENBERG, P. Emotional maturity as manifested in ideational interaction. *J. abnorm. soc. Psychol.*, 1955, **51**, 298–301.

LICHTENBERG, P., CASSETTA, RHONDDA K., & SCANLON, J. C. An analysis of types of treatment and an exploration of its utility. *J. nerv. ment. Dis.*, 1960, **130**, 353–365. (a)

LICHTENBERG, P., CASSETTA, RHONDDA K., & SCANLON, J. C. One description of mental health and disorder. *Arch. gen. Psychiat.*, 1960, **3**, 575–582. (b)

LICHTENBERG, P., KOHRMAN, R., & MACGREGOR, HELEN. *Motivation for child psychiatry treatment.* New York: Russell & Russell, 1960.

MASLOW, A. H. *Motivation and personality.* New York: Harpers, 1954.

SCOTT, W. A. Research definitions of mental health and mental illness. *Psychol. Bull.*, 1958, **55**, 29–45. (a)

SCOTT, W. A. Social psychological correlates of mental illness and mental health. *Psychol. Bull.*, 1958, **55**, 65–87. (b)

SZASZ, T. S. The myth of mental illness. *Amer. Psychologist*, 1960, **15**, 113–118.

(Received December 27, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 63, No. 3, 629–635

THE INFLUENCES OF PERSONALITY AND SITUATIONAL CAUTIOUSNESS ON CONCEPTUAL BEHAVIOR[1]

HOWARD A. MOSS

Child Research Branch, National Institute of Mental Health

A NUMBER of psychologists have stressed the importance of considering situational and personality factors as codeterminants of behavior. Newcomb (1943) dealt with the interaction of situational and personality factors in his classic study of attitude change among college students. As early as 1935 Lewin stated "the actual behavior of the child depends in every case both upon his individual characteristics and upon the momentary structure of the existing situation" (p. 71). In a recent review of the literature, Masling (1960) discusses the decreased predictive power of studies that fail to consider the combined influence of situational and personality factors.

The complex effects of personality and situational variables require more than the simple acknowledgment that they both contribute to behavior. Not all subjects are equally responsive to different situational cues. A situational variable may influence only the behavior of subjects who exhibit a particular personality characteristic. By pooling subjects, without regard for personality differences, the influence of the situation may be masked or greatly attenuated. On the other hand, the variance associated with the personality variable may be neutralized by strong situational cues. That is, different personality tendencies may be observable only under certain conditions. The optimum approach for assessing psychological processes should be based on the study of these two types of variables both as separate and combined determinants of behavior; as Rotter (1960) contends, it is necessary to study the interaction of situational and personality variables if we are to increase the power of our predictions and provide a basis for replicability.

This investigation deals with the way responses are modified by the subject's expectancy of success or failure on the experimental tasks. Persons who fear that they will be perceived in a negative light frequently exercise certain controls over their test behavior. These controls are defensive since they protect the person from revealing aspects of himself that might result in failure and/or disapproval. Defensive reactions should be greater when an individual has a generalized low expectancy for success or in a situation that specifically engenders low expectancies.

The term *cautiousness* is used to describe the class of defensive behaviors dealt with in this investigation. Cautiousness is regarded as "the tendency to behave in a manner designed to avoid potential failure or disapproval experiences, and this goal is achieved often at the expense of other satisfactions." That is, the cautious person is more concerned with avoiding failure than with obtaining success. In order for cautiousness to be observed there must be a basis for judging certain response alternatives as entailing less risk and less potential satisfaction than other alternatives. The level of aspiration paradigm provides a task situation that meets these criteria. The subject can behave in a cautious way by stating low goals, even when these expressed low goals place restrictions on his maximum possible score. That is, the subject is told that his score can be no higher than his expectancy statement. On the other hand, unrealistically high expectancy statements are discouraged by penalizing the subject for failing to score as high as he predicted he would.

The use of the level of aspiration as a measure of cautiousness has been discussed frequently in the psychological literature. Lewin, Dembo, Festinger, and Sears (1944) conclude that fear of failure is a major determinant of level of aspiration behavior.

[1] Adapted from a dissertation submitted to the graduate school of the Ohio State University in partial fulfillment of the requirements for the Doctor of Philosophy degree.

The author wishes to express his appreciation to Julian B. Rotter, committee chairman, for his invaluable assistance in the preparation of this report.

Persons who have a strong need to avoid failure behave cautiously by stating low goals. Frank (1935) states that cautiousness is a generalized personality trait that influences level of aspiration behavior. He did not obtain any independent measure of caution but inferred the trait from the level of aspiration behavior. Gould (1939) concludes from her results that goal setting behavior often serves a person as a protection of the ego in which the stated estimates are designed to shield him from failure.

Cautiousness is also reflected by the conceptual approach used in responses to visual stimuli. The verbal labels used to conceptualize a visual stimulus may vary considerably. For instance, responses to TAT pictures can be descriptive or, conversely, highly inferential. The descriptive statements are objective and their accuracy cannot be easily challenged. In addition, descriptive responses are nonprojective and provide the subject with a way of evading disclosure of internally held beliefs, conflicts, or needs. Descriptive responses are therefore assumed to represent a cautious approach since there is minimal risk of criticism and/or personal exposure. Inferential responses can be idiosyncratic, personalized, and imaginative. These responses are more open to evaluation of either a critical or favorable nature and would generally reflect a noncautious approach.

The hypotheses tested by this investigaare as follows:

1. Cautiousness, when induced by an experimental situation of failure, leads to increased descriptiveness in verbal responses to TAT pictures and conceptual sorts (MAPS figures).

2. The degree to which the failure situation leads to increased descriptiveness is determined by the subject's personality disposition. That is, personality and situational factors interact in the prediction of descriptive response.

3. Cautiousness, as a general personality characteristic (determined from level of aspiration behavior), is associated with greater descriptiveness in verbal responses to the personality test materials. The relationship should be most pronounced when assessed under neutral conditions (Control group).

METHOD

Subjects

The sample consisted of 109 students enrolled in the introductory psychology classes at the Ohio State University: 61 females and 48 males from 17 to 30 years of age.

Procedure

The subjects were randomly assigned to three experimental groups and each subject was seen individually for two separate sessions.

Session 1. Prior to the experimental conditions all subjects performed on the Rotter (1942) Level of Aspiration Board, to provide a behavioral measure of characteristic cautiousness.[2] The subjects were seated at the end of a table on which the Level of Aspiration Board was placed and the standard instructions were administered.[3] Each subject had 20 trials on this task, and a difference score (*D* score) was obtained between subject's actual performance score and his subsequent expectancy statement for each trial.

Rotter (1945) has described nine patterns characterizing level of aspiration behavior, regarded as "points of concentration which can be used as flexible standards for understanding individual cases," and based on a combination of three factors: his *D* score, the frequency of shifts, and the occurrence of unusual shifts (up after failure and down after success):

1. *Low positive D score pattern.* This is generally considered the most realistic response. Estimates are usually higher than past performances with an average number of shifts and unusual shifts absent.

2. *Low negative or very slightly positive D score patterns.* This pattern is similar to Number 1, but with a lower *D* score.

3. *Medium high D scores.* This pattern is characterized by frequent usual shifts (up after success and down after failure) and high *D* scores.

4. *Achievement followers.* These subjects' estimates are constantly changed to equal or closely approximate the previous performance.

5. *The step pattern.* Here the subject never lowers his estimate.

6. *Very high, positive D score patterns.* In this pattern the subject behaves unrealistically by stating high goals regardless of characteristically lower performance levels.

7. *High negative D score patterns.* This is characterized by frequent shifts down after success and negative *D* scores.

8. *Rigid pattern.* This pattern is defined by an absence of shifts.

9. *The confused or breakdown pattern.* Here there are a high number of shifts in addition to frequent un-

[2] A questionnaire was constructed as a second measure of characteristic cautiousness. This questionnaire proved to be a less effective predictor than the Level of Aspiration Board and was not included in the paper.

[3] Taken from unpublished manual prepared by J. B. Rotter.

usual shifts, i.e., shifts upward after failure and downward after success. *D* scores can be any size.

The subjects within each of the three experimental groups were classified as either cautious or noncautious from their level of aspiration pattern. Research findings indicate that the pooling of similar patterns provides greater generality and predictability than classifications based solely on the use of the specific patterns (Neems & Scodel, 1956). In this study, Patterns 1 and 3 were classified as noncautious and 2, 4, 7, and 8 as cautious. Ten subjects who obtained nonclassifiable patterns (5, 6, and 9) were omitted from the part of the analysis pertaining to personality differences. The interrater agreement for scoring the nine patterns was 85%; for the pooled cautious and noncautious patterns it was 97.5%.

Session 2. All subjects were administered a brief specially constructed paper-and-pencil form that they were told was a "test of social acceptability," consisting of 25 statements concerning social attitudes. The subjects were instructed to rate themselves on a five-point scale for each of these 25 statements. For Groups 1 and 2, the experimenter pretended to score the test in the presence of the subject and compared these results with false norms he had in front of him. The subjects in Group 1 were told that they scored in the "lower 10 percentile" (Failure group, $N = 36$),[4] and the subjects in Group 2 were told that they scored in the "upper 10 percentile" (Success group, $N = 36$). This procedure was omitted for Group 3 (Control group, $N = 37$). After experiencing the experimental condition all subjects were asked to take two more tests which they were told were also measures of "social acceptability." These two tests were to "make-up a story" for Pictures 2, 4, 17GF, and 20 of the TAT and to sort into two groups, according to to the subject's own idea, 22 cut-out human figures taken from the MAPS test (Shneidman, 1952). The sorting procedure was repeated three times for the complete group of figures. The TAT and MAPS responses were used to assess cautious conceptual behavior.

TAT. Each TAT story was scored for conceptual level on a four-point scale. The most descriptive stories obtained a score of 1 and the most inferential ones obtained a score of 4. Intermediate stories on the descriptive-inferential dimension and popular themes, as determined from normative studies (Eron, 1950; Rosenzweig & Fleming, 1949) were given scores of 2 and 3. The interrater reliabilities, based on product-moment correlations, for Pictures 2, 4, 17GF, and 20, respectively, were 1.00+, .67, 1.00+, and .95.[5]

MAPS Conceptual Sorting Task. The 22 figures in the MAPS test—including representations of males, females, figures with features, those where the features had been omitted, figures in costume, figures in everyday clothes, others in protective and gesticulating poses, and ones assuming a relatively static stance, etc.—suggested a variety of conceptual dimension. The subject had the opportunity to sort according to clearly physical qualities of the stimuli or on a more highly inferential basis. Each of the three MAPS sorts was scored according to the same 1-4 point scale of descriptiveness to inferentiality used in the scoring of the TAT stories. The interrater reliability for this score was 1.00+.[6]

Ohio State Psychological Examination (OSPE) scores were obtained from the university records for all the subjects for whom such scores were recorded. The OSPE scores provided an estimate of intellectual ability and were correlated with the scores obtained on the experimental measures.

TABLE 1
DESCRIPTIVE STATISTICS WITH RESPECT TO THE MAPS SORTS AND TAT CONCEPTUAL SCORES

	N	Sorts		TAT	
		M	*SD*	*M*	*SD*
Failure group					
Total	31	6.45	2.49	9.03	3.96
Cautious	14	6.50	2.56	9.07	4.28
Noncautious	17	6.41	2.43	9.00	3.68
Success group					
Total	32	7.00	2.44	9.38	3.40
Cautious	14	6.79	2.01	9.21	2.98
Noncautious	18	7.17	2.71	9.50	3.69
Control group					
Total	36	7.50	2.62	9.47	2.88
Cautious	15	6.13	2.00	8.33	2.55
Noncautious	21	8.48	2.58	10.29	2.83

RESULTS

The three groups (Failure, Success, and Control) were each dichotomized into cautious and noncautious subgroups on the basis of the subject's level of aspiration performance. The *N*, mean, and standard deviation for each group and subgroup are presented in Table 1. High scores on the sorts and TAT indicate inferential behavior and low scores reflect descriptiveness. The Mann-Whitney *U* test was used for comparing groups as to the degree of descriptive behavior on the MAPS sorts and TAT. The *U* values were converted into *z* scores[7] so that all group comparisons are expressed as *z* score values.

[4] At the conclusion of the experiment the Failure subjects were given information that negated the earlier failure experience.

[5] These coefficients were corrected, because a limited number of broad categories were used in rating these variables, by a procedure described by Peters and Van Voorhis (1940). Lack of precision in the correction results in some correlations of 1.00+. These correlations do indicate reliabilities in the high .90s.

[6] See Footnote 5.

[7] Siegel (1956) states that as the *N* in either group increases in size, the sampling distribution of *U* approaches the normal distribution. Therefore, a formula

TABLE 2
DIFFERENCES BETWEEN EXPERIMENTAL GROUPS ON DESCRIPTIVENESS
(expressed as z score values)

Condition	Sorts	TAT
Failure vs. Success	.91 *F*[a]	.31 *F*
Failure vs. Control	1.45 *F*	.87 *F*
Success vs. Control	.68 *S*	.61 *S*

Note.—z score of 1.96 needed for .05 level.
[a] The letter *F*, *S*, or *C* indicates whether Failure, Success, or Control group was higher on descriptiveness.

TABLE 3
DIFFERENCES BETWEEN EXPERIMENTAL GROUPS (Cautious and Noncautious subgroups analyzed separately) ON DESCRIPTIVENESS
(expressed as z score values)

Condition	Noncautious subjects		Cautious subjects	
	Sorts	TAT	Sorts	TAT
Failure vs. Success	.76 *F*[a]	1.56 *F*	.49 *F*	.39 *F*
Failure vs. Control	2.41*** *F*	1.27 *F*	.59 *C*	.07 *C*
Success vs. Control	1.50 *S*	.88 *S*	1.10 *C*	.86 *C*

[a] The letter *F*, *S*, or *C* indicates whether Failure, Success or Control group was higher on descriptiveness.
*** $p < .02$; two-tailed test.

Table 2 presents data pertaining to the first hypothesis. While the Failure group was more descriptive than the Success or Control groups for all comparisons, as hypothesized, these differences were not statistically significant. Thus, the hypothesis pertaining to the influence of conditions was not supported.

The second hypothesis stated that the Failure condition would have its greatest effect in increasing descriptive responses when the subject's personality disposition was taken into consideration. This was tested by comparing the influence of the conditions for the cautious and the noncautious subjects separately. The findings pertaining to this hypothesis are reported in Table 3. As can be seen, the Failure condition promoted increased descriptive responses for noncautious but not for cautious subjects. However, the only statistically significant comparison was between the Failure and Control conditions for the MAPS sorts ($z = 2.41$, $p < .02$; two-tailed). The cautious subjects did not respond differentially to Failure as compared with Control and Success conditions. The increase in descriptiveness under the Failure condition, noted in Table 2, appears, thus, to have been contributed by the noncautious subjects.

The mean scores presented in Table 1 provide a basis for summarizing and better understanding the foregoing results. The cautious subgroup in each of the three experimental conditions were quite similar in descriptiveness. They had attained preexperimentally, a plateau on descriptiveness and were not appreciably influenced by the conditions. The noncautious subjects were most inferential in the Control condition and became more descriptive in the experimental conditions, particularly in the Failure condition. Therefore, the cautious subjects remained stable on descriptiveness across conditions, whereas, the noncautious subjects exhibited a degree of descriptiveness comparable to the cautious subjects only in the Failure condition.

According to the final hypothesis, the cautious subjects should be more descriptive in their responses to the sorts and TAT than noncautious subjects, and the difference should be most pronounced for a neutral situation (Control condition). Table 4 shows the comparisons between the cautious and noncautious subgroups for descriptiveness on the sorts and TAT within each condition.

TABLE 4
DIFFERENCES BETWEEN CAUTIOUS AND NONCAUTIOUS SUBGROUPS ON DESCRIPTIVENESS
(expressed as z score values)

Condition	Sorts	TAT
Failure (Cautious vs. Noncautious)	.01 NC[a]	.06 NC
Success (Cautious vs. Noncautious)	.27 C	.10 C
Control (Cautious vs. Noncautious)	2.84**** C	2.07** C

[a] NC or C indicates whether the noncautious or cautious subgroup was higher on descriptiveness.
** $p < .05$; two-tailed test.
**** $p < .01$; two-tailed test.

is provided for converting the obtained U into a z score when the N in either group exceeds 20. This conversion is also recommended as a correction when the same rank occurs for scores involving both groups. Since all of the observations meet either one or both of these conditions, all of the U's are converted into z scores.

TABLE 5

PRODUCT-MOMENT CORRELATIONS BETWEEN EXPERIMENTAL MEASURES AND OSPE SCORES

Condition	Sorts	TAT
Failure group ($N = 28$)	−.02	−.10
Success group ($N = 30$)	.32*	−.02
Control group ($N = 28$)	.02	−.52****

* $p < .10$; two-tailed test.
**** $p < .01$; two-tailed test.

The z score was again used to evaluate the significance of the difference for these comparisons.

The cautious and noncautious subgroups did not differ in descriptive behavior for the Failure and Success condition. However, for the Control condition, the cautious subjects were significantly more descriptive than the noncautious subjects on both the sorts ($z = 2.84$, $p < .01$; two-tailed) and TAT ($z = 2.07$, $p < .05$; two-tailed). The effects of the personality variable appear to have been obscured by the strong situational forces operating for the Failure and Success conditions. When these situational forces were minimized (Control condition) the influence of the personality variable became quite evident. While the cautious subjects were not more descriptive in all conditions as predicted, the fact that they demonstrated greater descriptiveness in the neutral situation indicated a basic relationship between descriptiveness and cautiousness. That is, descriptiveness on the sorts and TAT was associated with characteristic cautiousness and this relationship held only for the Control condition.

Scores on the OSPE were available for the majority of the subjects. Correlations with the conceptual scores on the TAT and MAPS sorting task for each of the situational conditions are presented in Table 5.

The MAPS conceptual score and OSPE were correlated for the subjects in the Success condition ($r = .32$, $p < .10$; two-tailed). These MAPS responses were confounded with an experimental condition; this correlation is, therefore, difficult to interpret. On the other hand, the TAT conceptual score and OSPE were negatively related for the Control condition ($r = -.52$, $p < .01$; two-tailed). That is, the brighter subjects gave more descriptive TAT responses.

The two significant correlations reported in Table 5 increase the complexity of the experimental findings. These relationships do not appear to detract from the stated hypotheses, but rather, suggest that additional dimensions also contributed to the conceptual behavior.

The TAT and MAPS conceptual scores were positively correlated ($r = .30$, $p < .01$; two-tailed). Cautiousness on the Level of Aspiration and OSPE scores were unrelated ($r_{bis} = .16$, not significant).

DISCUSSION

When all subjects were pooled there was no relation between descriptiveness and the situation. However, when personality differences were considered, the situation became a significant determinant of descriptive behavior, only noncautious subjects being influenced by experimental conditions. Failure induced descriptiveness in the noncautious subjects, but did not increase descriptive responses among the cautious subjects, who had a pre-experimental set to respond to all conditions as if failure were imminent. This finding highlights the importance of systematically investigating individual differences in susceptibility to situational cues. By so doing, one is able to increase the power of predictive hypotheses. This position is underscored by some findings obtained by Marlowe and Crowne (1961), who observed that persons differ in their need to subscribe to socially desirable behaviors and that differences in this need predict sensitivity to situational cues.

The second major finding was that personality differences predicted descriptive behavior only for a neutral situation. That is, cautious subjects were more descriptive than noncautious subjects for the Control condition, but not under Failure or Success, which apparently masked the variance otherwise associated with the personality variable. This finding bears on the need to consider both situational and personality variables. Certain situational conditions can easily obscure personality differences leading investigators to conclude erroneously that no differences exist.

Fear of failure is a function of both situ-

ational and personality factors. Subjects defended against it by being more descriptive on the conceptual tasks. On both the MAPS and TAT, tests frequently used to assess content variables, the personality and situational variables were both associated with descriptiveness. Yet the possible occurrence of many substantive responses is contingent on the conceptual level used by the subject. If a subject is highly descriptive, the range of potential responses is reduced. Furthermore, pronounced descriptiveness precludes the occurrence of many personalized responses which, by definition, are inferential. These findings emphasize the importance of considering the presence of general defensive patterns, such as cautiousness, in interpreting psychological test results.

No specific predictions were made concerning the Success group, which turned out to be intermediate between the Failure and Control groups on descriptive behavior. The positive reinforcement given the Success group might have been expected to decrease Cautious behavior. However, such was not the case. An alternative explanation, in line with the obtained findings, is that the positive reinforcement on the social attitudes test actually evoked cautiousness on the subsequent, functionally related tasks. That is, the provision of a reinforcement for one task (positive or negative) may structure other tasks in that situation as potential success or failure experiences. The reinforcement, regardless of its direction, increases involvement for the related tasks. Support for this position is suggested by Lazarus, Baker, Broverman, and Mayer (1957), who state that "psychological stress involves the thwarting of a motive state" and that "a motive state has to be present in order for stress to occur." Furthermore, in keeping with the results of this study, they found that "stressor conditions" served to make subjects low in achievement motivation more literal on a task where they were required to listen to and reproduce tape recorded paragraphs.

In addition to showing the interplay between situational and personality variables, the study provides evidence for the moderate generality of cautiousness over a variety of situations. Cautiousness on the Level of Aspiration task predicted descriptiveness on the sorts ($r_{bis} = .56$, $p < .01$) and on the TAT ($r_{bis} = .42$, $p < .05$) for the Control condition. Descriptiveness on the sorts and TAT were positively correlated for the total sample. Furthermore, cautiousness induced by the Failure situation predicted descriptive behavior.

SUMMARY

This investigation dealt with the influence of situational and characteristic cautiousness as codeterminants of conceptual behavior. A Failure condition produced situational cautiousness and the level of aspiration measured characteristic cautiousness. Cautious conceptual responses consisted of descriptive responses to TAT cards and a figure-sorting task.

The major findings were as follows:

1. The Failure condition contributed a significantly greater amount of descriptive responses than the Control condition only for subjects judged to be noncautious prior to the experimental conditions. When all subjects were pooled this relationship largely disappeared.

2. The personality variable predicted descriptive behavior only for a neutral situation. That is, cautious subjects were significantly more descriptive than noncautious subjects on the sorts and TAT for the Control condition, but not for the Failure or Success conditions.

REFERENCES

ERON, L. D. A normative study of the Thematic Apperception Test. *Psychol. Monogr.*, 1950, **64**(9, Whole No. 315).

FRANK, J. D. Some psychological determinants of the level of aspiration. *Amer. J. Psychol.*, 1935, **47**, 285–293.

GOULD, ROSALIND. An experimental analysis of level of aspiration. *Genet. psychol. Monogr.*, 1939, **31**, 3–115.

LAZARUS, R. S., BAKER, R. W., BROVERMAN, D. M., & MAYER, J. Personality and psychological stress. *J. Pers.*, 1957, **25**, 559–577.

LEWIN, K. *A dynamic theory of personality*. New York and London: McGraw-Hill, 1935.

LEWIN, K., DEMBO, TAMARA, FESTINGER, L., & SEARS, PAULINE S. Level of aspiration. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944. Pp. 333–378.

MARLOWE, D., & CROWNE, D. P. Social desirability

and response to perceived situational demands. *J. consult. Psychol.*, 1961, **25**, 109–115.

MASLING, J. The influence of situational and interpersonal variables in projective testing. *Psychol. Bull.*, 1960, **57**, 65–85.

NEEMS, R., & SCODEL, A. Authoritarianism and levels of aspiration scores. *J. soc. Psychol.*, 1956, **43**, 209–215.

NEWCOMB, T. M. *Personality and social change.* New York: Dryden, 1943.

PETERS, C. C., & VAN VOORHIS, W. R. *Statistical procedures and their mathematical bases.* New York: McGraw-Hill, 1940. Pp. 393–399.

ROSENZWEIG, S., & FLEMING, E. E. Apperception norms for the Thematic Apperception Test. *J. Pers.*, 1949, **17**, 475–503.

ROTTER, J. B. Level of aspiration as a method of studying personality: II. Development and evaluation of a controlled method. *J. exp. Psychol.*, 1942, **31**, 410–422.

ROTTER, J. B. Level of aspiration as a method of studying personality: IV. The analysis of patterns of response. *J. soc. Psychol.*, 1945, **21**, 159–177.

ROTTER, J. B. Some implications of a social learning theory for the prediction of goal directed behavior from testing procedures. *Psychol. Rev.*, 1960, **67**, 301–316.

SHNEIDMAN, E. S. Manual for the Make-a-Picture-Story method. *Proj. tech. Monogr.*, 1952, No. 2, 1–92.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956. Pp. 120–126.

(Received December 28, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 63, No. 3, 636-638

CRITIQUE AND NOTES

DOGMATISM AND ACQUIESCENT RESPONSE SET

EDWARD LICHTENSTEIN, ROBERT P. QUINN
University of Michigan

AND GERALD L. HOVER[1]
Veterans Administration Hospital, Ann Arbor, Michigan

THE Dogmatism scale has been one of the major instruments used by Rokeach and his associates in their studies of belief systems. According to Rokeach (1960), "the primary purpose of this scale is to measure individual differences in openness or closedness of belief systems. Because of the way we have defined open and closed . . . the scale should also serve to measure general authoritarianism and general intolerance" (pp. 71–72). Rokeach has criticized the California F Scale on the grounds that the F Scale is not independent of ideological preference and therefore measures "right" authoritarianism and not general authoritarianism. For example, communists score low on the F Scale but high on Dogmatism, or *general* authoritarianism. The Dogmatism and F scales usually correlate highly, however, the reported correlations ranging from .54 to .77 (Rokeach, 1956).

The Dogmatism and F scales are structurally identical, both being Likert-type scales. Items on both scales are positively scored attitudinal statements, a high number of "Agree" responses yielding a high score. On both scales the subject responds by endorsing items with numbers ranging from +3 to −3, 0 excluded, indicating responses ranging from strong agreement to strong disagreement.

The construct validity of the F Scale has been criticized because several investigators have shown that part of the scale's variance is attributable to an acquiescent response bias (Chapman & Campbell, 1957; Christie, Havel, & Seidenberg, 1958; Cohn, 1953; Jackson & Messick, 1957; Leavitt, Hax, & Roche, 1955). Some writers have even felt impelled to reinterpret the meaning of authoritarianism so as to account for the data on F Scale response set (Gage, Leavitt, & Stone, 1957).

In spite of the structural similarity of the Dogmatism and F scales, the susceptibility of Dogmatism to acquiescent response set has not yet been questioned. The present study was designed to demonstrate that Dogmatism is just as vulnerable to acquiescent response set as the F Scale.

METHOD

Subjects were 40 adult male inpatients in the neuropsychiatric service of a General, Medical, and Surgical Veterans Administration Hospital. The subjects' mean age at nearest birthday was 33.7 years (SD = 9.8). Their mean number of years of formal education was 9.9 (SD = 1.7). On the seven-point occupational status rating scale described by Warner, Meeker, and Eells (1949), the subjects' occupations had a mean value of 5.8 (SD = 1.0); 27 subjects were unskilled or semiskilled laborers. In religious affiliation 32 subjects were Protestant and 8 were Roman Catholic. All subjects were tested within 4 weeks of their admission to the neuropsychiatric wards.

The F Scale, Form 40-45, was given under conventional instructions (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950). The split-half reliability estimated by the Spearman-Brown formula was .79 for our sample (M = 4.5; SD = 0.9). The 40-item Dogmatism scale was given under instructions recommended by its author (Rokeach, 1956). Split-half reliability was .76 (M = 4.1; SD = 0.7).

Two measures of acquiescent response set were developed. The first was a Difference score based on 32 items taken from a measure of impulsivity developed by Sanford, Webster, and Freedman (1957). A revised form of Sanford's scale was given to subjects with Likert-type response categories. Sanford's measure contained 121 positively and 23 negatively scored items. From this item pool 16 negatively scored items were selected for which 16 positively scored items, similar to the negatively scored ones in content, could be found. For example, a typical positive item, "I have used alcohol excessively," was paired with a corresponding negative item, "I have never done any heavy drinking." For each subject two scores were obtained, a total score on the 16 positive items and a total score on the 16 corresponding negative items. Subtracting the negative from the positive total yielded a Difference score for each subject. The size of this Difference score was assumed to reflect the magnitude of each subject's acquiescent bias. The split-half reliability of the measure was .62 (M = −7.7; SD = 16.8). Chapman and Campbell (1957) and Couch and Keniston (1960) have also used content canceled items to measure acquiescence.

[1] The authors wish to express their appreciation to the members of the neuropsychiatric staff of the Ann Arbor Veterans Administration Hospital who assisted in the collection and processing of the data. All statements are those of the authors and do not necessarily represent the opinions or policy of the Veterans Administration.

TABLE 1
INTERCORRELATIONS AMONG ACQUIESCENCE MEASURES, THE F AND DOGMATISM SCALES AND VERBAL INTELLIGENCE

	2	3	4	5
1. Acquiescence difference score	.29*	.47**	.45**	−.26
2. Acquiescence MMPI True		.51**	.38**	−.15
3. Dogmatism scale			.70**	−.20
4. F Scale				−.32*
5. Shipley-Hartford (verbal intelligence)				

* Indicates significance at the .05 level, one-tailed test.
** Indicates significance at the .01 level, one-tailed test.

The second acquiescence measure was total number of MMPI True responses; its reliability was .93 (M = 234.6; SD = 41.0). This measure was analogous to those used by Mathews (1929), Lorge (1937), Cronbach (1942), Cohn (1953), Shelley (1956), and Gage et al. (1957).

In the first testing session subjects were given in fixed succession: several pencil-and-paper tests being used for other purposes, the Shipley-Hartford, the F Scale, Dogmatism, and the Sanford measure from which Difference scores were obtained. The vocabulary section of the Shipley-Hartford was used as a measure of verbal intelligence. Its raw-score mean was 25.2 (SD = 6.0). In the second testing session, a week later, subjects were given the MMPI in card form but were not permitted to use the Uncertain response category.

RESULTS

The Pearson product-moment correlations among the measures are presented in Table 1. The obtained correlations were in the expected direction. Both acquiescence measures correlated significantly with both Dogmatism and the F Scale. The −.32 correlation between the F Scale and the Shipley-Hartford supports the earlier reports of similar negative correlations between the F Scale and intelligence by Adorno et al. (1950), Cohn (1951), and Gough (1951). The correlations between the acquiescence measures and the Shipley-Hartford are comparable to the low negative relationship between acquiescence and intelligence reported by Cohn (1953) and Bass (1956). The previously reported findings just cited were largely based on student samples; the present study extends the generality of these findings to a sample of neuropsychiatric patients both older and less educated than the student groups. The only puzzling aspect of the data was the relatively low .29 correlation between the two acquiescence measures.

Another way of interpreting the number of MMPI True items is as a measure of degree of psychoticism rather than acquiescence. This interpretation would follow from Cottle and Powell's (1951) report that endorsing all MMPI items True gives a high score on the MMPI "psychotic triad" while endorsing all items False elevates scores on the "neurotic triad." Thus the obtained relationships between MMPI True responses and the F and Dogmatism scales could mean that subjects who scored high on the F Scale and Dogmatism tend to be the more psychotic. Two considerations tend to rule out this explanation. First is the fact that the pattern of correlations between True MMPI responses and the other variables was quite similar to the pattern based on these variables and the Difference score. Second are some additional data that explicitly test this interpretation. The final hospital diagnosis was used to classify subjects into four groups: psychotic, neurotic, personality disorder, and psychophysiological reaction. A chi square test showed no association between diagnostic category and either the F Scale or Dogmatism. The mean F Scale and Dogmatism scores were slightly *higher* for the 15 neurotics than for the 11 psychotics; the mean number of MMPI True responses was also slightly higher for the neurotics. Neither of these differences was significant, however. All these results are inconsistent with the alternative explanation that number of MMPI True responses reflects degree of psychoticism and strengthens the assumption that this is a valid measure of acquiescent response set.

DISCUSSION

These data indicate that Rokeach's Dogmatism scale is vulnerable to acquiescent response set roughly to the same extent as the F Scale. This finding suggests that previous work with Dogmatism should be re-examined with an eye to determining which measures are related to Dogmatism by virtue of their sharing variance in acquiescence. As with the F Scale, the correlation between Dogmatism and other response biased measures cannot be interpreted unambiguously.

Future studies relating the F Scale and Dogmatism to other possibly acquiescence-biased measures should include an independent measure of acquiescence as a matter of routine. This will make it possible to specify to what extent obtained relationships are attributable to construct rather than methods variance. For example, in the present study the correlation of .70 between the F Scale and Dogmatism overstates the relationship between the two measures due to content alone. When the effects of the two acquiescence measures are partialed out this correlation reduces to .57.

The recent attempt by Couch and Keniston

(1960) to characterize the "personality" of the consistently acquiescent respondent fails to recognize the acquiescence-vulnerability of Dogmatism. Because their measure of acquiescence correlates .40 with Dogmatism, they conclude that this "appears to reflect the diffuse anxiety of the yeasayer, and to suggest that he seeks dependency on external figures for support" (p. 159). On the other hand, their data indicate only that the acquiescent respondent is characterized by a high score on a personality measure with an acquiescence bias built into it. The hypothesis that the Dogmatism scale is contaminated by acquiescence bias not only offers a more parsimonious interpretation of Couch and Keniston's results but also receives substantial support in the present study.

SUMMARY

The study was designed to test the hypothesis that Rokeach's Dogmatism scale, structurally identical to the F Scale, is similarly vulnerable to acquiescent response bias. Two measures of acquiescence were developed and administered, along with the Dogmatism and F scales and a measure of verbal intelligence, to 40 adult neuropsychiatric inpatients. Both acquiescence measures were positively related to both the Dogmatism and F scales. It was concluded that the Dogmatism measure contains acquiescent response set variance. The implications of this finding for past and present research using the Dogmatism scale were discussed.

REFERENCES

ADORNO, T., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D., & SANFORD, R. *The authoritarian personality.* New York: Harper, 1950.

BASS, B. Development and evaluation of a scale for measuring social acquiescence. *J. abnorm. soc. Psychol.*, 1956, **53**, 296–299.

CHAPMAN, L., & CAMPBELL, D. Response set in the F Scale. *J. abnorm. soc. Psychol.*, 1957, **54**, 129–132.

CHRISTIE, R., HAVEL, JOAN, & SEIDENBERG, B. Is the F Scale irreversible? *J. abnorm. soc. Psychol.*, 1958, **56**, 143–159.

COHN, T. Some intellective and non-intellective factors in grade-getting. Paper read before Western Psychological Association, San Jose, California, April 27, 1951.

COHN, T. Factors related to scores on the F Scale. Unpublished doctoral dissertation, University of Michigan, 1953.

COTTLE, W., & POWELL, J. The effect of random answers to the MMPI. *Educ. psychol. Measmt.*, 1951, **11**, 224–227.

COUCH, A., & KENISTON, K. Yeasayers and naysayers: Agreeing response set as a personality variable. *J. abnorm. soc. Psychol.*, 1960, **60**, 151–174.

CRONBACH, L. Studies of acquiescence as a factor in the true-false test. *J. educ. Psychol.*, 1942, **33**, 401–415.

GAGE, N., LEAVITT, G., & STONE, G. The psychological meaning of acquiescence set for authoritarianism. *J. abnorm. soc. Psychol.*, 1957, **55**, 98–103.

GOUGH, H. Studies of social intolerance: I. Some psychological and social correlates of anti-Semitism. *J. abnorm. soc. Psychol.*, 1951, **33**, 237–246.

JACKSON, D., & MESSICK, S. A note on ethnocentrism and acquiescent response sets. *J. abnorm. soc. Psychol.*, 1957, **54**, 132–134.

LEAVITT, H., HAX, H., & ROCHE, J. "Authoritarianism" and agreement with things authoritative. *J. Psychol.*, 1955, **40**, 215–221.

LORGE, I. Gen-like: Halo or reality? *Psychol. Bull.*, 1937, **34**, 545–546. (Abstract)

MATHEWS, C. The effect of the order of printed response on an interest questionnaire. *J. educ. Psychol.*, 1929, **20**, 128–134.

ROKEACH, M. Political and religious dogmatism: An alternative to the authoritarian personality. *Psychol. Monogr.*, 1956, **70**(18, Whole No. 425).

ROKEACH, M. *The open and closed mind.* New York: Basic Books, 1960.

SANFORD, N., WEBSTER, H., & FREEDMAN, M. Impulse expression as a variable of personality. *Psychol. Monogr.*, 1957, **71**(11, Whole No. 440).

SHELLEY, H. Response set and the California attitude scales. *Educ. psychol. Measmt.*, 1956, **16**, 63–67.

WARNER, W., MEEKER, MARCHIA, & EELLS, K. *Social class in America.* Chicago: Science Research Associates, 1949.

(Received October 31, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 63, No. 3, 639-642

THE EFFECTS OF INFORMATION SOURCE STATUS AND DOGMATISM UPON CONFORMITY BEHAVIOR

ROBERT N. VIDULICH AND IVAN P. KAIMAN
Louisiana State University

THE present investigation attempted to provide a partial answer to the question: are highly authoritarian persons uniformly conforming to all directive-producing sources, or are they differentially influenced by information sources that are high or low in status? Previous research has indicated that both variables, source status and authoritarianism are related to conformity behavior (e.g., Crutchfield, 1955; Lefkowitz, Blake, & Mouton, 1955).

Rokeach (1960) in his theory of general authoritarianism or "dogmatism," differentiates between persons who are relatively *open* (low dogmatic or low authoritarian) or *closed* (high dogmatic or high authoritarian) with respect to their belief-disbelief systems about reality. A person's belief-disbelief system is relatively open or closed depending on his ability to "receive, evaluate and act on relevant information received from the outside on its own intrinsic merits . . . the more closed the belief-system, the more difficult should it be to distinguish between information received about the world and information received about the source [of this information] . . ." (pp. 57–58).

Rokeach's (1960) theory would predict that relatively closed persons should confuse what an external authority had to say about a situation with who the authority was, leading to increased conformity to the authority if he were high in status, and decreased conformity to, or even rejection of, the authority were he seen as being low in status. Relatively open individuals should be capable of discriminating source from message, and should show little difference in their reactions to varying source figures.

METHOD

Apparatus

The experimental situation used the autokinetic effect and the general procedure of Sherif (1935). Subjects in the present study were asked to report the *direction* rather than the distance of autokinetic movement.

The stimulus light box was mounted on a table 32 inches above the floor at one end of a lightproof and semisoundproof room. A point of light was exposed through a hole 1 millimeter in diameter in one end of a lightproof wooden box, 9 inches long, 3 inches wide, and 3 inches high, painted flat black inside and out. The light source was a small radio bulb at the far end of the box. The bulb was connected to a transformer to receive a constant electrical input of 2.5 volts. Two thicknesses of tissue paper diffused the light. Exposure time was automatically controlled by two Hunter timers so that the light was alternately activated and shut off for 10 second intervals. This sequence was repeated for the total experimental period. The light was visible at all times except when the experimenter switched off the power supply between the two phases of the experiment.

The subjects and the experimenter were seated at the opposite end of the experimental room from the light source, the subject being 197 inches (5 meters) from the light aperture. The experimental room was in complete darkness at all times.

Subjects

From a pool of 307 undergraduates in introductory psychology courses at Louisiana State University who had taken the 40-item Dogmatism scale, Form E (Rokeach, 1960), 30 high and 30 low scorers were selected as subjects. All 60 subjects were in the upper or lower quartiles of Dogmatism scale scores, all were females between the ages of 18 and 23 years, and all but 4 were sophomores. None had previous knowledge about the autokinetic phenomenon. The mean Dogmatism scale score for the 30 open subjects was 126.75 (σ = 11.66); for the 30 closed subjects, the mean Dogmatism score was 181.37 (σ = 11.87).

Each subject was in the experimental situation with a confederate of the experimenter throughout the experiment. The 60 subjects were divided into four groups: Group I consisted of 15 subjects who had scored high on the Dogmatism scale (closed), for whom the confederate was identified as a high status person; Group II consisted of 15 closed subjects for whom the confederate was identified as a low status person. Groups III and IV each consisted of 15 subjects who had scored low on the Dogmatism scale (open), for whom the confederate was identified as being high or low in status, respectively.

Procedure

During Phase A of the experiment, the subject observed 30 exposures of the light, judging the direction of the light's "movement," either left or right, and privately recording these judgments on one or another of two 3 × 5 pads in front of him. During Phase B, the subject observed 30 additional exposures of the light, now making his responses verbally, with the experimenter recording these responses. The confederate reported his "judgments" prior to those of the subject during Phase B.

Eighty percent of the confederate's responses during Phase B were randomly given in the direction least frequently reported by the subject in Phase A. Upon determining the subjects least predominant directional choice of Phase A, the experimenter informed the

TABLE 1
TREATMENT OF DIFFERENCE SCORES

Analysis of Variance

Source	*SS*	*df*	*F*	*p*
Source status	56.06	1	3.95	*ns*
Dogmatism	19.26	1	1.36	*ns*
Interaction (S × D)	308.29	1	21.74	<.01
Within	794.12	56		
Total	1177.73			

Comparison of Subgroup Means

Dogmatism	Source status: High		Source status: Low		*t*	*p*
	M[a]	*SD*	*M*[a]	*SD*		
Closed group	5.93	4.45	−0.53	4.53	3.29	<.01[b]
Open group	2.53	2.69	5.13	3.02	2.48	<.05
t	2.52		4.02			
p	<.01[b]		<.01[b]			

[a] $N = 15$ for each group.
[b] One-tailed test of significance used for these comparisons.

confederate of the response pattern to be reported by tapping him on the right or left shoulder. The confederate used a string of beads and knots to determine the direction of each response to be given.

The subject entered the already darkened room in which the confederate was seated, and was guided to his seat by the experimenter. The meeting of the subject and the confederate took place in complete darkness. For Groups I and III (high status source), the confederate was presented as "Dr. John Wilson, who has been doing a good deal of research at LSU. He will be observing the same thing as you and we are interested in comparing your reactions with his." For Groups II and IV (low status source), the confederate was presented as "John Wilson, who is a high school student at Baton Rouge High. He will be observing the same thing as you and we are interested in comparing your reactions with his." These introductory statements were made to the subject while the experimenter was escorting him to the experimental room.

Task instructions were given as follows:

> When I give the signal "Ready" a light will go on. After a short time, the light will start to move. Then the light will go off. When the light goes off you are to mark the direction of the light's movement on one of the two pads directly in front of you. There is a pencil between the two pads. If the light moved to the left, make a mark on the left pad; if the light moved to the right, make a mark on the right pad. Sometimes the light will move a great distance, at other times it will move only slightly. Be sure you make a mark for each trial, and *only one* mark to indicate the direction in which the light moved on each trial. When I give the signal there will be *no* talking. Ready.

The 30 trials of Phase A were then administered. After the thirtieth exposure, the experimenter said:

> Please hand me your pads. In order to speed this up a little bit, the next time the light goes on, I want you to tell me which way the light moved, and I'll record your answers myself. This should take less time. You can tell me your answers after the light goes off. And out of courtesy to the faculty [visiting student], Dr. Wilson [John] you can go first.

At the termination of Phase B, the subject was asked not to discuss the procedures of the experiment with anyone. The experimenter then escorted the subject out of the room, asking the confederate to remain "to talk to you for a few minutes." Thus the subject never saw the confederate.

RESULTS

The first response measure defining conformity was calculated by subtracting the number of the least dominant directional responses of Phase A from the number of same direction responses during Phase B for each subject (*difference score*). The second response measure was calculated by tabulating the number of responses of each subject which agreed with those of the confederate during Phase B (*agreement score*). For both measures, the higher the score, the greater the presumed conformity to the confederate's responses.

From the upper portion of Table 1, it is evident that neither dogmatism nor status source is significantly related to the shift in directional judgments from Phase A to Phase B (difference score). The

TABLE 2
TREATMENT OF AGREEMENT SCORES

Analysis of Variance

Source	*SS*	*df*	*F*	*p*
Source status	355.26	1	30.73	<.01
Dogmatism	96.26	1	8.33	<.01
Interaction (S × D)	589.09	1	50.96	<.01
Within	647.32	56		
Total	1687.93	59		

Comparison of Subgroup Means

Dogmatism	Source status: High		Source status: Low		*t*	*p*
	M[a]	*SD*	*M*[a]	*SD*		
Closed group	18.27	3.89	7.13	2.45	9.36	<.01[b]
Open group	14.53	3.88	15.93	3.15	1.09	*ns*
t	2.62		8.80			
p	<.01[b]		<.01[b]			

[a] $N = 15$ for each group.
[b] One-tailed test of significance used for these comparisons.

major source of variance is attributable to the interaction of these two variables ($F = 21.74$, $p < .01$). The lower portion of Table 1 indicates that the closed group exposed to a high status source shifted most toward the confederate's responses from Phase A to Phase B (difference = 5.93), while the closed-low status group shifted *away* from the "judgments" of the confederate (difference = −0.53). The open groups displayed dissimilar behavior, agreeing significantly more with the directional responses of the low status than the high status source. That these results are not due to differential responsiveness during Phase A is indicated by the following: the mean number of least dominant directional responses during Phase A are 11.40, 10.80, 10.87, and 10.93 for Groups I through IV, respectively.

The analyses of agreement scores generally corroborate the foregoing results. Analysis of variance (upper portion of Table 2) indicates that both source status and dogmatism are highly significant sources of variation when considered alone, a finding which differs from the difference score analysis, but that the interaction of these two variables again produces the greatest effect ($F = 50.96$, $p < .01$).

Again, the closed subjects agreed more with the high than with the low status source (lower portion of Table 2). Out of 30 possible agreements during Phase B, the closed groups averaged 18.27 agreements under the high status condition and only 7.13 agreements when responding following the low status source ($t = 9.36$, $p < .01$). The open groups did not differ significantly in mean agreement responses under the two source status conditions.

DISCUSSION

The results of this experiment generally support the contention that conformity is a function of both the status of the source of information to which the person is exposed, and the authoritarian tendencies of the person being influenced. Analyses of both conformity measures used indicate that a highly important interaction exists between the variables of general authoritarianism (cognitive closedness) and information source status, which is more influential in the determination of conformity behavior than either variable taken alone. These data amply illustrate the importance of jointly considering field conditions and person characteristics in examining conformity/influence behavior.

That these results cannot be attributed to intelligence and/or cognizance of the "rigged" nature of the situation is supported by two ancillary findings. No differences were found between open and closed subjects in gross intellectual ability (as estimated by academic grade point average). Previous studies with the Dogmatism scale also indicate that it is unrelated to intelligence for this population (Rokeach, 1960). With regard to the second possibility, no subject reported having suspected that the autokinetic movement of the light was not real, or that the confederate was anything but what he was introduced as.

With regard to the postulated inability of the closed person to discriminate a message from its source, the present findings are seen as suggestive rather than definitive. What is required in future studies of this hypothesis is the use of an experimental situation in which it is possible for the subject to evaluate independently the information received about reality from a source and the source itself. A modified Asch-type procedure, such as that used by Crutchfield (1955), would appear to be appropriate.

Two weaknesses of the present study are readily apparent. The control conditions of having the subjects alone in Phase B, and having the subjects in the presence of the confederate without having the confederate make responses during Phase B, would be helpful in determining the reasons for the presently-puzzling behavior of the open groups which tended to conform more to low than to high status sources. Also, the reliability of the two response measures used in this study has not been established. A third experimental phase in which the subject's judgments are not preceded by those of the confederate is required to determine if it is the *judgment* of the subject or his *announcement* of his judgment that is altered in the present situation. Similar attempts to determine the temporal duration of "conformity" are notably lacking in previously-reported studies in this area also.

SUMMARY

This study tested the prediction that cognitively open (high authoritarian) and closed (low authoritarian) persons should be influenced differentially by high and low status information sources.

Groups of 30 open and closed subjects, selected on the basis of extreme scores on Rokeach's Dogmatism scale, were placed in an autokinetic situation. Each subject privately judged the directional movement of the light for 30 exposures and then made an additional 30 judgments following reports of a confederate of the experimenter. For one-half of each group, the confederate was presented as a college professor (high status); he was introduced as a high school student for the low status groups. During the second experimental phase, the confederate made responses which were

80% in the least-judged direction of the subject during the private judging experimental phase.

Data analyses for two response measures of conformity fully support the prediction of an interaction between cognitive closedness (authoritarianism) and information source status. The closed groups agreed significantly more with the high status than with the low status source. Open groups tended to agree more with low than with high status sources.

REFERENCES

CRUTCHFIELD, R. S. Conformity and character. *Amer. Psychologist*, 1955, **10**, 191–198.

LEFKOWITZ, M., BLAKE, R. R., & MOUTON, JANE S. Status factors in pedestrian violation of traffic signals. *J. abnorm. soc. Psychol.*, 1955, **51**, 704–706.

ROKEACH, M. *The open and closed mind*. New York: Basic Books, 1960.

SHERIF, M. A study of some social factors in perception. *Arch. Psychol.*, 1935, No. 187.

(Received October 31, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 63, No. 3, 643-645

MANIFEST CONTENT OF DREAMS OF CONVICTED SEX OFFENDERS[1]

MARK I. GOLDHIRSH

State Correctional Institution, Graterford, Pennsylvania

KARPMAN'S (1954) extensive review of the literature on sex offenders and Ramsey's (1953) critical review of dreaming show the total absence of systematic quantitative investigations of dream material from sex offenders. During the past 6 years of administering therapy to convicted institutionalized sex offenders, I have observed that their dreams had a relatively high frequency of sexual content and uncamouflaged unlawful sex activity. The present study is an attempt to test these observations objectively and systematically. The two previous systematic investigations into the dreams of convicts (Hanks, 1940; Selling, 1932), failed to include a control group and the sex criminal group was not studied.

METHOD

The subjects were 16 therapy cases selected from my files. All the inmates were confined at the State Correctional Institution at Graterford, Pennsylvania, and had volunteered for therapy. The data were collected between 1954 and 1960.

In order to subject the hypothesis of this study to a quantitative systematic evaluation, a scoring manual was developed. Several hundred dreams of inmates with records of sexual offenses were studied and compared to dreams of convicted nonsexual offenders. Material in the dreams that seemed to differentiate the two groups provided the basis for expanding and refining the scoring categories.

A brief version of the scoring system used to identify sexual themes in the manifest content of the dreams follows.[2]

The dream text will score positive (+) "sexual" if it contains words, slang expressions, activity, or references to the following items:

1. "Wet dreams" or any form of sex play.
2. Sexual organs or their activity (erection, masturbation).
3. Nakedness, or body description of any erotic nature: Sexy, lovely, beautiful, or well-built breasts, hips, thighs, legs, nipples, body, etc.
4. Kissing, hugging, petting, necking, or erotic feelings from body contact.
5. Sex crime: Any form of sex play with a child or teenager until 16 years of age. Any activity involving incest, rape, homosexuality, or sodomy (oral or anal genital contacts). An uncamouflaged "sex crime" dream also receives a double score (++).

The following are examples of dreams that are scored "sexual" (+):

1. Woman in her early forties came into my room. She had large *breasts*, full face, dark eyes, short hair, and a *negligee* on. She just killed someone. We had *coitus* and enjoyed it. I had this same dream before but couldn't see the whole body and did not have *sex* before. She has the physical qualities of my mother.

2. A girl about 45 years old—she is *good looking* and I ask her out. We went to a movie, ate, and *I kissed her* and got an *erection*.

3. I had a *wet dream* with a girl. This is an improvement. She was 30 years old.

4. I was in some sort of camp or gathering where persons were dispossessed by a flood and had to sleep where they could. We were all jammed together sleeping. I went to sleep and awoke. The person next to me was touching me and it was a very nice girl. She was asleep. It was a shock to me. I felt around and *made advances*. I finally *got down to business* but there was no response. She did not awaken. Just as I was *reaching a climax* I awoke. I had a *wet bed*.

5. I was on the beach swimming with a girl in the *nude*. She had long red hair. I don't like red. I didn't recognize her. I had a portable radio and plenty to drink. We swam and lay in the sun but there was no *sex* as I can remember.

6. This woman suggested that we take a hot shower. So we went upstairs, *undressed* and we both went into the shower. We started *fooling around*, *washed each other* on the back, and *I laid her* there in the bathtub. As I was *coming* in the dream *I came* over the bed.

The following are examples of dreams that are scored uncamouflaged "sex crime" and they receive a double score (++):

1. My father shakes my hand and says that they forgive me for having *sex with boys*. I couldn't forgive myself.

2. *A little boy plays with my penis and squeezes* it. I do not get a hardon. There is a glass on my belly and I am supposed to get a *hardon*.

3. Lately over and over mostly *sex dreams* and they never finish. I am trying to have *sex with a young girl about 12 years old*. It is hard to see her face. *I undress her* and *play around with her tits* and *love her up*. But I never get to finish up. Looks like the first girl *I got in trouble with sexually*. My wife is in the corner sneering at me.

4. *I had boys play with my dick*. Then I went home and cried.

5. *I am trying to get my prick into a boy*. I have a

[1] This study was supported in part by Research Grant 534 from the National Institute of Mental Health, United States Public Health Service.

I am indebted to Aaron T. Beck for his guidance and helpful suggestions.

[2] The unabridged scoring manual has been deposited with the American Documentation Institute. Order Document No. 6862 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

TABLE 1
MATCHING DATA ON SEX CRIMINAL (SC) AND OTHER CRIMINAL (OC) PATIENTS

Pair	Age		Ethnic group		IQ		Marital status		Crime pattern[a]		Age first crime		Socioeconomic level		Nature of crime	
	SC	OC	SC	OC	SC	OC	SC	OC	SC	OC	SC	OC	SC	OC	SC	OC
A	54	50	W	W	109	116	M	M	Yes	Yes	42	32	Middle	Low	Sodomy[b] (child[c])	Robbery
B	53	50	W	W	113	115	S	S	Yes	No	48	48	Middle	Middle	Sodomy (teens[d])	Murder
C	40	43	W	N	112	102	S	S	Yes	Yes	12	16	Low	Low	Sodomy (teens)	Burglary
D	42	34	W	W	112	104	S	S	Yes	Yes	24	19	High	Middle	Sodomy (child)	Robbery
E	25	32	W	W	101	84	M	M	Yes	Yes	15	23	Low	Low	Sodomy (child)	Burglary
F	44	32	W	W	98	114	S	S	Yes	No	39	30	Low	Low	Sodomy (child)	Larceny
G	26	28	W	W	107	105	S	S	Yes	Yes	23	18	Low	Low	Sodomy (child)	Burglary
H	33	33	W	W	104	110	M	M	No	No	32	29	Middle	Middle	Sodomy (child)	Robbery

[a] Crime pattern means more than one conviction for current type of crime or an admitted extensive history.
[b] Sodomy is defined as sexual activity involving oral or anal relations.
[c] Child means until 9 years old.
[d] Teens is 10–16 years of age.

TABLE 2
FREQUENCY OF "SEXUAL" AND UNCAMOUFLAGED "SEX CRIME" DREAMS OF MATCHED SEX OFFENDER AND OTHER CRIMINAL PATIENTS

Pair	Number of sexual dreams out of 10**		Number of uncamouflaged sex crime dreams out of 10*	
	SC	OC	SC	OC
A	10	1	7	0
B	5	2	3	0
C	5	0	0	0
D	4	0	2	0
E	9	0	9	0
F	3	1	1	0
G	2	0	1	0
H	4	0	1	0

* $p = .01$, one-tailed, Wilcoxon matched-pairs signed-ranks test.
** $p = .005$.

hardon and can't get it in. I awoke with a *hardon.* I was ashamed and afraid to tell you.

6. I am a tree surgeon and a *kid is playing with my penis.*

After the scoring system was developed, it was then applied to a sex criminal group and a control group. The sex criminal group consisted of eight therapy cases with an extensive pattern of dangerous criminal sexual behavior against children. They were matched patient for patient as closely as possible with a control group (no arrests for sex crimes and no admitted sexually deviated history) on the basis of age, race, intelligence, marital study, pattern to their type of crime, age at time of first offense, and socioeconomic level. The relevant identifying data on all patients are listed in Table 1. The dreams of the eight matched pairs of subjects were not part of the series of dreams used in the construction of the scoring manual.

The first 10 dreams reported during the therapy process were abstracted from each case record and typed on individual sheets of paper. The total sample of 160 dreams (10 per patient for 16 patients) was then arranged in random order and presented to another rater[3] who was unfamiliar with the proposed research. After a few minutes of clarification she then scored the dreams using only the scoring system of five steps presented in this paper. She had no prior knowledge concerning these cases or their dreams. These blind ratings were subjected to statistical evaluation reported in Table 2. In order to get an estimate of the reliability of the ratings, I also rated the dreams.

RESULTS AND DISCUSSION

The raters agreed as to the presence or absence of a sexual element on 158 out of 160 dreams. This is almost 99% agreement. The scoring procedure, thus, is highly reliable. The results of the comparison between the two groups is given in Table 2.

There is virtually no overlap between the groups: 53% of the dreams of the sex criminal patients contain one or more sexual elements, whereas in the other criminal group, 5% of their dreams contain one or more sexual elements.

[3] The author wishes to thank Stephanie Manning for her scoring help.

The raters further agreed as to the presence or absence of an uncamouflaged element of sex crime on 156 out of 160 dreams. This is almost 98% agreement. Note that in every case except one, the sexual offender had one or more dreams containing an uncamouflaged sex crime theme (Table 2), whereas with the eight convicts in the control group there was no reference to sexual criminal activity in any of their 80 dreams.

On the basis of these results, the conclusion seems warranted, within limits set by the institutionalized population and the small sample, that male sexual offenders show a greater incidence of sexual and sex crime dreams than do other types of convicts.

REFERENCES

HANKS, L. M. An explanation of the content of dreams through an interpretation of dreams of convicts. *J. gen. Psychol.*, 1940, **23**, 31–46.

KARPMAN, B. *The sexual offender and his offenses.* New York: Julian, 1954.

RAMSEY, G. V. Studies of dreaming. *Psychol. Bull.*, 1953, **50**, 432–440.

SELLING, L. S. Effect of conscious wish upon dream content. *J. abnorm. soc. Psychol.*, 1932, **27**, 172–178.

(Received November 14, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 63, No. 3, 646-649

SELF-ESTEEM, MANIFEST HOSTILITY, AND EXPRESSION OF HOSTILITY

MILTON E. ROSENBAUM AND ROBERT F. STANNERS
State University of Iowa

A number of recent studies have demonstrated complex relationships between personality test measures and aggressive behavior. The findings have suggested that the test responses may reflect individual differences both in level of potential expression and inhibition of aggressive response. Two such measures are the Self-Esteem (SE) scale and the Siegel (1956) Manifest Hostility Scale (MHS). Two studies (Rosenbaum & deCharms, 1960, in press) employing the SE scale have suggested that under conditions of threat to self-esteem, subjects with low SE scores are more strongly instigated to aggression than subjects who have high SE scores but that inhibition of expression of aggression in social situations is greater among low SE subjects. Two other studies (Berkowitz, 1960; Hokanson & Gordon, 1958) have found that under conditions of low arousal to aggression subjects with high MHS scores express more aggression than low MHS subjects but under conditions of high arousal, subjects low in MHS increase their level of expression of aggression to a greater extent than subjects with high MHS scores.

In the previous studies relating SE scale scores and hostile response, control conditions in which aggression arousal was low were not employed. It was not possible, therefore, to examine the character of aggressive expression prior to instigation and consequent inhibition. The present study was in part designed to correct this fault. In addition, it has been suggested (Rosenbaum & deCharms, in press) that a decrease of the inhibition in expression of aggression may occur when the aggression is measured in a nonsocial context, i.e., in response to TAT cards as in the present study. Finally, the present circumstance allows for replication of the previous studies relating MHS scores to aggressive response but in a slightly altered context. In the present study, the subject is subjected to personal attack in contrast to the indirect arousal procedures used in the other studies.

METHOD

Students in the elementary psychology course were administered a variety of personality measures, including the Self-Esteem scale (Rosenbaum & deCharms, 1960) and Siegel's (1956) Manifest Hostility Scale. Based on their scores on the SE scale, male subjects were recruited in pairs to participate in an experiment on "creativity."

In the laboratory, subjects participated in a procedure adapted from a portion of Weiss and Fine's (1956) study. All subjects were asked to complete three easy (1–3) designs from the Arthur Stencil Design Test I following the standard instructions. Subjects in the Arousal condition then were given five difficult designs (16–20) with a time limit of one minute for each design (no subject completed this task satisfactorily). Between each design, the experimenter made increasingly derogatory comments concerning the subject's performance. Following the three easy designs subjects in the Nonarousal condition were given five relatively easy designs (4–8) with a one-minute time limit (all completed the designs satisfactorily). Between each design, the experimenter praised the subject's performance.

All subjects then were presented with two cards from the TAT series (2 and 8BM) for which they were to write stories for 5 minutes each. The experimenter continued to offer derogatory or favorable remarks to the Arousal and Nonarousal subjects, respectively. Consistent with the procedure employed by Hokanson and Gordon (1958), Card 2 may be regarded as a Low Cue for aggression and Card 8BM as a High Cue. The order of presentation of the cards was counterbalanced within the Arousal and Nonarousal groups. Finally, subjects responded to a questionnaire containing five questions of which Number 5 requested the subject to indicate his impression of the experimenter on a six-point scale ranging from "very favorable" to "very unfavorable."

Forty male subjects participated in the experiment. One-half of the subjects were exposed to the Arousal procedure and one-half to the Nonarousal treatment. In addition, one-half of the subjects in each of these conditions had low SE scale scores (45 or below) and the other half had high SE scale scores (50 or above). No effort was made to select subjects on the basis of MHS scores.

RESULTS

Responses to Question 5 of the post-experiment questionnaire, which assessed the favorability of the subjects' impression of the experimenter, may serve to evaluate the effectiveness of the Arousal manipulation. The analysis of variance of the scores derived from Question 5 is summarized in Table 1. The results indicate that the procedures produced differential hostility toward the experimenter between the Arousal and Nonarousal subjects. The obtained means were 3.0 and 1.6, respectively, a high score representing a negative evaluation. There was no significant difference between High and Low SE subjects, nor was there a significant SE $\times$ Arousal interaction.

The TAT stories were scored by the category system employed by Hokanson and Gordon

TABLE 1

ANALYSIS OF VARIANCE OF THE FAVORABILITY OF IMPRESSIONS OF THE EXPERIMENTER

Source	*df*	*MS*	*F*
Arousal	1	20.30	18.13**
Self-Esteem	1	1.80	1.61
Arousal × Self-Esteem	1	.07	—
Error (within)	36	1.12	

** $p = < .01$.

TABLE 2

ANALYSIS OF VARIANCE OF TAT SCORES AS A FUNCTION OF LEVEL OF SE AND AROUSAL CONDITION

Source	*df*	*MS*	*F*
Self-Esteem	1	5.00	1.07
Arousal	1	14.45	3.09
Self-Esteem × Arousal	1	2.45	—
Error (between)	36	4.67	
Cue	1	24.20	16.46**
Cue × Self-Esteem	1	.80	—
Cue × Arousal	1	6.05	4.12*
Cue × Self-Esteem × Arousal	1	.45	—
Error (within)	36	1.47	
Total	79		

* $p = < .05$.
** $p = < .01$.

(1958).[1] The principal rater was not informed of the conditions in which the subjects appeared and his reliability was checked by partial scoring on the part of the experimenter. Table 2 presents the summary of the analysis of variance of these scores. Both the main effect for Cue and the interaction of Cue × Arousal are significant. These results reflect greater aggressive response to the High Cue than the Low Cue card. In addition, the subjects were differentially responsive to the Arousal conditions on only the High Cue card in that subjects in the Arousal condition presented greater aggression to the High Cue card than did subjects in the Nonarousal condition. Both findings principally support the experimental manipulations. The results fail to indicate any differential effect of level of SE.

Subjects in each Arousal condition were split into two groups on the basis of the over-all median of the MHS scores. Table 3 presents the resulting *N*s and mean TAT hostility scores. The previous analysis (Table 1) indicated a highly significant difference between the two levels of cue. In order to facilitate analysis of variance by the approximation method suggested by Walker and Lev (1953)

[1] We are grateful to Jesse E. Gordon for providing the scoring guide.

TABLE 3

MEAN TAT HOSTILITY SCORES AND MULTIPLE RANGE TESTS FOR EACH CUE LEVEL

	Arousal		Nonarousal	
	High MHS (*N* = 8)	Low MHS (*N* = 12)	High MHS (*N* = 11)	Low MHS (*N* = 9)
High cue*	5.00[a]	2.50[b]	2.09[b]	2.11[b]
Low cue*	2.38[c]	1.50[d]	1.45[d]	1.67[d]

* Means *within* each row marked with a letter in common do not differ significantly at the .05 level by multiple range test (Kramer, 1956).

TABLE 4

ANALYSES OF VARIANCE OF TAT SCORES AS A FUNCTION OF LEVEL OF MHS AND AROUSAL CONDITION

Source	High Cue			Low Cue		
	df	*MS*	*F*	*df*	*MS*	*F*
Manifest hostility	1	2.719	8.31**	1	.142	—
Arousal	1	1.538	4.70*	1	.107	—
Manifest hostility × Arousal	1	1.588	4.86*	1	.299	—
Error	36	0.327		36	.4551	

* $p = <.05$.
** $p = < .01$.

separate analyses of the results for each level of cue were conducted and are presented in Table 4. The significant Arousal × MHS interaction for the High Cue scores is clarified by reference to the results of the multiple range test (Kramer, 1956) presented in the upper row of Table 3. A significant difference is demonstrated between high MHS and low MHS subjects under Arousal while no difference is indicated between levels of MHS under Nonarousal.

The Low Cue results did not show a significant Arousal × MHS interaction although the trend of the means is in the same direction as found for the High Cue responses. As support for the trends indicated in the Low Cue findings, the multiple range test[2] presents significant differences between the mean for the high MHS subjects in the Arousal condition and each of the other means but no difference among these latter means (lower row, Table 3).

A complex interaction among SE, MHS, and Arousal is possible but not readily assessed. Statistical evaluation is precluded by excessively imbalanced cell entries when all three variables are included. Inspection of means does not suggest statistical interaction of the variables.

[2] The multiple range test does not necessarily require a significant *F* ratio.

Scores on both the SE scale and the MHS were available from the source population in the elementary psychology course. Product-moment correlations were computed. For males ($N = 120$), $r = -.173$ which is not significant at .05 level. A low but significant ($p = < .01$) negative relationship is found for females ($r = -.244$, $N = 169$).

DISCUSSION

The present results present no support for a relationship between SE scale scores and hostility expression in TAT stories despite apparent success in experimental arousal of aggression. The present procedure differs from earlier studies in which SE scores have been found to relate to hostile response on the dependent measure of hostility. In previous studies aroused subjects were permitted to respond directly to the source of instigation either orally or by means of written personality sketches. In the present study, only indirect imaginative responses were assessed. In contrast with previous studies in which arousal was accomplished by a peer in a clearly arbitrary manner, the present arousal procedure was carried out by the experimenter (a graduate student) and was task related. This difference may have contributed to excessive inhibition, thereby obscuring arousal differences between levels of SE.

Interpretation of the results that indicate a direct relationship between Arousal and MHS scores in their effect on hostility expression does not require reference to inhibitory concepts. In previous studies differences in the direction of greater hostility on the part of subjects with high MHS scores appeared under Nonarousal conditions and a greater increase in hostility expression occurred for low MHS subjects under arousal conditions. This result has been interpreted as reflecting the operation of an inhibitory variable for the aroused high MHS subjects based on the production of strong aggression-anxiety at high levels of aggression potential. In the present situation, MHS level is related to hostile responses only in the Arousal condition and the results indicate a positive relationship; the higher the MHS score the higher is the hostile response in the TAT stories. As indicated, this relationship must be regarded only as a trend for the Low Cue condition.

Several possible explanations are available to account for the discrepancies in results obtained in the present and previous studies employing the MHS scores. The Arousal procedures differ in that the subject is directly derogated by the experimenter in the present study while in earlier studies more indirect methods of arousal were employed. Hokanson and Gordon (1958) relied on the presentation of the case of a juvenile delinquent who was portrayed as more or less sadistic in order to effect High and Low Arousal, respectively. Berkowitz's (1960) subjects received messages from a peer indicating either liking or disliking of the university and its students to create differential arousal. One possibility is that the present procedure served to arouse stronger aggression although, as has been suggested, the character of the instigator and the task relevance of the attack is likely to have produced greater inhibition. To reconcile these factors it may be suggested that the heightened strength of instigation may have served to raise the goal gradient for hostile response so that in relative terms, the influence of the inhibition gradient was not apparent in the results. High MHS subjects in particular could proceed "closer" to the goal of aggression without crossing the inhibitory gradient.

Both earlier studies found greater hostility expressed by high MHS subjects under Low Arousal conditions which is in contrast to the present study in which no difference is obtained between the MHS levels. The present procedure emphasized approbation throughout the subjects performance while it is possible that at least to some degree hostility was aroused in the Low Arousal condition in the earlier studies. It may be suggested that some minimal degree of arousal is necessary before differentiation in hostile expression occurs for the different MHS levels.[3]

SUMMARY

The relationship of two personality measures to hostile response under conditions of differential arousal to hostility was evaluated. Responses to TAT cards served as the dependent measure of hostility. The arousal procedures produced differential response to the TAT cards but the measure of level of self-esteem was not related to hostile expression. Following arousal, subjects with high scores on the Manifest Hostility Scale (MHS) responded with greater hostility than low MHS subjects to the TAT cards. No difference in response was found between aroused and nonaroused subjects with low MHS scores. Discrepancies between the present results and those obtained in other experiments were discussed primarily in terms of differences in procedure.

REFERENCES

BERKOWITZ, L. Manifest hostility level and hostile behavior. *J. soc. Psychol.*, 1960, **52**, 165–171.

[3] In an unpublished study, Cizak found no significant relationship between MHS scores and TAT responses obtained under Nonarousal conditions from Naval Reserve subjects.

HOKANSON, J. E., & GORDON, J. E. The expression and inhibition of hostility in imaginative and overt behavior. *J. abnorm. soc. Psychol.*, 1958, **57**, 327–333.

KRAMER, C. Y. Extension of multiple range tests to group means with unequal numbers of replications. *Biometrics*, 1956, **12**, 307–310.

ROSENBAUM, M. E., & DECHARMS, R. Direct and vicarious reduction of hostility. *J. abnorm. soc. Psychol.*, 1960, **60**, 105–111.

ROSENBAUM, M. E., & DECHARMS, R. Self-esteem and the expression of aggression. In N. Washburne (Ed.), *Decisions, models and groups*. Vol. II. New York: Pergamon, in press.

SIEGEL, S. M. The relationship of hostility to authoritarianism. *J. abnorm. soc. Psychol.*, 1956, **52**, 368–372.

WALKER, HELEN M., & LEV, I. *Statistical inference*. New York: Holt, 1953.

WEISS, W., & FINE, B. J. The effect of induced aggressiveness on opinion change. *J. abnorm. soc. Psychol.*, 1956, **52**, 109–115.

(Received November 14, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 63, No. 3, 650–653

A TEST OF THE PROPOSITION: WE WANT TO BE ESTEEMED MOST BY THOSE WE ESTEEM MOST HIGHLY

CECIL R. WURSTER, BERNARD M. BASS, AND WAYNE ALCOCK
Louisiana State University

ESTEEM is the value ascribed to a person, as a person, regardless of the position that he occupies (Bass, 1960, p. 277). A tenet of modern personality theory that is basic to several schools of psychotherapy, seemingly axiomatic because of its widespread acceptance in psychology, is that we are motivated to maintain or increase our own esteem in the eyes of others. "This cherished possession [selfhood] must forever be made more adequate, more worthy, and it must forever be defended against stain and injury, whether from the acts of others or of the valuing organism" (Murphy, 1947, p. 523).

The maintenance and enhancement of esteem is directly revealed in our employing such defense mechanisms as rationalization, identification, or projection, as well as in our tendency to overestimate our esteem more often than we underestimate it (Bass, 1960, pp. 324–325). The desire to support and enhance one's own esteem suggests a dynamic relationship significant for leadership theory: we want to be esteemed most by those we esteem the most.

The proposition can be derived as follows:

I want to be valued (p. 364).[1] Some people, in my eyes, are more able or more valuable or more estimable than others (p. 277). I accept more readily and attach more weight to the influence and opinions of those I esteem most highly (p. 289). I want to be valued by those whom I value most highly.

It follows also that:

I will want those I esteem the most to know of my worthy performances. Therefore, I will be motivated to influence those most whom I most esteem about matters enhancing my own esteem.

What devolves here is a circular leadership process. "I esteem you. You can influence me. I will want to influence you to look upon me favorably. I will be less likely to attempt to influence those whom I do not esteem, to regard me favorably." This devolution, in turn, is consistent with the proposition that "we are more likely to interact, the more we esteem or value each other" (p. 359).

Thus, for example, we would infer that the classroom instructor should be much more concerned to have his efforts favorably viewed by those whom he regards as good students than by those he regards as poor students. In turn students should be more concerned about their evaluation from an instructor whom they regard highly than from an instructor whose opinions they discount.

This study tests the proposition that we are motivated to influence most those whom we esteem the most. Esteem is based on ability to some extent (p. 284), and since a person's abilities vary from one situation to another, say from a scholastic to a social situation, so should his esteem. At the same time, esteem appears to transfer through stimulus generalization across situations (p. 287). A second purpose of the study was therefore to see whether the proposition is bounded by specific situations. That is, the study sought to examine which of the following two propositions is more adequate:

I want to be esteemed in a particular situation by those I esteem specifically in that situation.

I want to be esteemed in a variety of situations by those whom I esteem regardless of the situation.

METHOD

Esteem: Scholastic and Social

One hundred ten students from several elementary psychology classes were asked to list the first names of 15 campus friends. The 110 students listed a total of 1,650 friends. No doubt, the same person was cited by several respondents. Indeed, the names of many of the respondents are likely themselves to have appeared on the lists of other respondents. But our attention was focused on 110 independent respondents, each with their 15 independent responses, not on any objective attributes of those identified in the 1,650 responses.

Each respondent was asked to rank his 15 friends in the order he would prefer each "as a study partner." This rank order of friends provided an index of the extent each friend was *esteemed scholastically* by the respondent. The actual esteem accorded to a particular person by all his associates was never obtained, only what one respondent thought of him.

In the same way, each respondent was asked to rank his friends again according to the extent he would prefer each "to join a social club of mine." This rank order provided an index of the *social esteem* of each friend mentioned.

The instructions were worded, thus:

. . . Assume that you're to recommend the people that are included in your list for membership in a

[1] All subsequent page references are from Bass (1960).

social club of which you're a member. You're to recommend them in your order of preference for membership. Place the number *1* after the name of the person you would *most* want to be in your club. [Experimenter pause] Place the number *15* after the name of the person you would *least* want to be in your club. [Experimenter pause] Now rank all other names in your list. Be sure to use all numbers between *1* and *15*. Don't use any number more than once . . .

Now assume that you're to recommend people that are included in your list for an honorary award for scholastic achievement. You're to recommend these people in your order of preference for this award. Place the number *1* after the name of the person you would recommend first. [Experimenter pause] Place the number *15* after the name of the person you would recommend last. [Experimenter pause] Now rank all other names in your list. Be sure to use all numbers between *1* and *15*. Do not use any number more than once.

Targets Selected or Avoided for Communications about Oneself

Each respondent was given a list of 15 questions in the scholastic area such as:

Which friend would you most want to know that you have a superior IQ on an intelligence test?

Which friend would you least want to know that you have had to take a remedial arithmetic course?

Which friend would you most like to know that you have won a scholarship?

Intermixed among these scholastic questions were 15 *socially* relevant questions such as:

Which friend would you most want to know that you have dated a campus personality?

Which friend would you tell first that you have a new sports car?

Which friend would you least want to know that you do not have the proper clothes to wear to a social function?

Finally, 15 buffer questions that overlapped both areas or were occasionally irrelevant were likewise scrambled in the total list of 45 questions:

Which friend's opinion do you value the most?

Which friend would you least like to know that you have a religious problem?

Which friend would you least like to know that you have sprained your ankle?

RESULTS AND DISCUSSION

Relation between Esteem and Selection as Target of Communication

The same single friend could be mentioned by a respondent as much as 15 times on the 15 social items or on the 15 scholastic items. But only 3 respondents did this of the 110. A Poisson distribution of frequency of repeated mentions emerged as shown in Table 1. Over half of the acquaintances were *never* mentioned by respondents in a given social or scholastic area on *any* of the 15 items. On the other hand, extremely few friends earned more than nine mentions from a respondent in either area. It appeared easy for respondents to list 15 acquaintances by first name and to rank them. Yet most respondents had a smaller cadre of close friends to whom they preferred to direct all favorable information about themselves.[2] Figure 1 shows how this preference centered on those friends they most esteemed. Thus, friends ranked first in esteem socially received an average of 4.0 mentions in response to the 15 social questions. Similarly, friends ranked first scholastically received 3.7 mentions of the 15 scholastic items. A sharply accelerating curve appears to describe the relation between the social esteem with which a respondent holds a friend and his tendency to single out the friend as a preferred recipient of favorable social information. The same holds for scholastic esteem and selection as a recipient of scholastic information.

TABLE 1

FREQUENCY THAT A FRIEND WAS MENTIONED ON 15 SOCIAL QUESTIONS AND ON 15 SCHOLASTIC QUESTIONS

Frequency with which the same friend is mentioned in response to the 15 questions	Number of friends mentioned on:	
	15 social questions	15 scholastic questions
0	895	1006
1	372	288
2	178	151
3	78	65
4	59	40
5	26	26
6	19	24
7	9	19
8	3	11
9 or more	11	19
	1650	1650

An unexpected and possibly significant reversal appeared. Note in Figure 1 that minimum mention as a preferred recipient occurs for those ranked 11th and 12th in esteem rather than 14th or 15th. Some kind of perversity in responding particularly in the social area seemed to develop among about 10% of the respondents. For example, no respondent gave any friend ranked 11th, 12th, or 13th more than three mentions. Yet those friends ranked 14th or 15th in esteem received four mentions as targets for favorable communication from four respondents and five mentions from two respondents.

[2] Wherever we state "desired recipient of favorable information," we also imply "*undesired* recipient of *unfavorable* information." The 15-item lists were balanced almost equally with respect to favorable and unfavorable information.

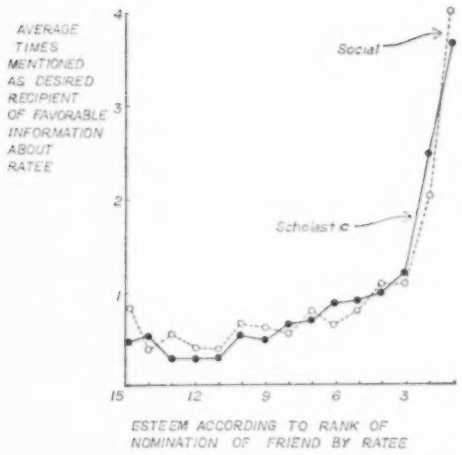

FIG. 1. Relation between esteem of friend according to rank of nomination and average times mentioned as desired recipient of favorable information.

A few subjects may possibly have misinterpreted instructions and ranked as "15" their most esteemed friend and as "1" their least esteemed friend. Yet this "mechanical" failure seems most unlikely in view of the specific instructions that were read word-for-word by the experimenter.

In designating the preferred recipient of favorable information about the respondent, one mentioned a bottom-ranker in his own esteem hierarchy 7 times, another 9 times, and one even 12 times! The latter was an extreme number of mentions even for top ranking friends. It seems doubtful that chance or haphazard responding is involved, in these bizarre results for no similar extreme numbers of mentions emerged anywhere else in the scatterplot other than here or for the most highly esteemed members. As many as 10% of the respondents may, thus, single out as recipients of favorable information those they esteem least perhaps as a means of self-enhancement through increased contempt or contrast. A suggested dynamic for these respondents might be: "I value myself. I do not value you. I tell you how good I am. This furthers my feeling how much better I am than you by contrast."

General or Specific Relationship?

As can be seen in Figure 1, the curvilinear relations between esteem and mentions were the same in both the social and scholastic areas. But the question remains as to whether halo or general rating bias was responsible, or whether social esteem was associated specifically with social rather than scholastic mentions and scholastic esteem with scholastic rather than social mentions.

The data were made linear by collapsing the 15 esteem ranks into two classes: the upper 7 and the lower 8 ranks, and by collapsing into two classes the frequency with which a friend was mentioned: mentioned at least once vs. never mentioned. Phi coefficients were then computed for each of the 110 subjects between the tendency for the subject to place a friend in the upper or lower half in esteem and the tendency to mention the friend as a desired recipient of favorable information. In collapsing these cells we undoubtedly reduced the strength of the associations displayed in Figure 1, but at this point we were interested in examining the relative strength of associations by situation. Table 2 shows the average phi coefficients, corrected for the collapse of 15 into two classes, to provide estimates of product-moment correlations that would have been obtained had we not dichotomized the data. The standard error of the mean correlations based on the distributions of 110 values are also given showing that the distance from the obtained means and zero were about 10 times their standard errors.

The comparable mean correlation between the tendency to be selected as the recipient of social and scholastic information was .44 while the mean rank-difference correlation between the 15 social esteem rankings and the 15 scholastic esteem rankings assigned by each of the 110 respondents was .39.

Social esteem appeared to make a friend equally desirable as the recipient of either socially ($\bar{\phi}$ = .41) or scholastically ($\bar{\phi}$ = .42) favorable information about oneself. On the other hand, scholastic esteem seemed somewhat more specifically associated with desirability as a target for favorable scholastic communications about oneself ($\bar{\phi}$ = .43) than favorable social communications about oneself ($\bar{\phi}$ = .22).

These results suggest that if we admire someone socially we are likely to want him to know of our academic and social successes. But if we esteem someone scholastically, we tend to be more con-

TABLE 2

CORRECTED PHI COEFFICIENTS BETWEEN SOCIAL AND SCHOLASTIC ESTEEM AND TENDENCY TO BE SELECTED AS THE DESIRED RECIPIENT OF INFORMATION ABOUT THE RATEE

Desired as recipient of favorable information	Esteem	
	Social	Scholastic
Social	.43 ± .041	.22 ± .041
Scholastic	.42 ± .036	.43 ± .038

cerned that he learn of our academic worth than our social reputation. The specific situations were selected, of course, not for their own sake but to test a proposition and they may not be equipotent. On a college campus, particularly in the post-Sputnik era, scholastic success may enhance social self-esteem but social success may have less effect on scholastic appraisal.

Implications

Extrapolation from these results suggests that we tend to behave in a manner directed at maintaining and enhancing our own esteem *in the eyes of those we esteem* rather than in the opinion of just anybody. That we are, thus, more responsive to the suggestions of those we esteem buttresses the *coercive* effects of esteem on success as a leader. One of us (Bass, 1960) has noted elsewhere that esteem can produce coercive as well as persuasive success as a leader, although the possible coerciveness arising from one's power as a person (irrespective of position) has not been much appreciated or investigated (p. 289). We follow the esteemed person because we believe him to have the ability and the power as a person to reward or punish us. Part of this coercive power, these data suggest, derives from the effects of the esteemed figure in granting or denying to us words of support, encouragement, and admiration.

REFERENCES

BASS, B. M. *Leadership, psychology and organizational behavior*. New York: Harper, 1960.

MURPHY, G. *Personality*. New York: Harper, 1947.

(Received November 21, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 63, No. 3, 654–655

A CROSS-CULTURAL INVESTIGATION OF THE RELATIONSHIP BETWEEN ETHNIC PREJUDICE, AUTHORITARIAN IDEOLOGY, AND PERSONALITY[1]

ARON WOLFE SIEGMAN[2]

Bar-Ilan University, Israel

THE major finding reported in *The Authoritarian Personality* (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950) is that prejudice is part of a larger attitudinal or ideological syndrome, which can be measured by the F Scale, and which in turn is associated with certain personality needs and defense mechanisms, e.g., aggression, repression, projection, etc. Although serious questions have been raised about the adequacy of the original sample (Hyman & Sheatsley, 1954), subsequent studies with diverse groups have produced essentially similar results (Christie, 1954, p. 155). In the United States, however, ethnic tolerance plays a prominent role in the ideologies of leftist and liberal political groups. Thus in selecting subjects who vary on the prejudice variable one may automatically select along a political radicalism-conservatism dimension. This raises the possibility that the obtained correlates of ethnic prejudice may really be related to the radicalism-conservatism dimension rather than to prejudice as such.

The present study of the relationship between ethnic prejudice, authoritarian ideology, and personality in a group of Israeli college students included two measures of ethnic prejudice. One measure was designed to evaluate the subjects' prejudice against Arabs, and the other the subjects' prejudice against recent Jewish North African immigrants. Whereas political parties in Israel differ in their attitudes towards the Arabs—the leftist parties advocating greater equality for the Arabs—prejudice against the North African immigrants is equally condemned by all political parties. Thus the present study afforded the opportunity of investigating the correlates of ethnic prejudice that is independent of political commitment.

METHOD

Subjects. The subjects consisted of 38 female and 16 male undergraduate psychology majors at Bar-Ilan University. One half of the group was European-born and the other half Israeli-born, but all subjects were of European middle class background. The subjects' age range was from 19 to 24 with about 85% of the subjects in the age range 19–22.

Authoritarian ideology scale. The 10 F Scale items[3] with the highest discriminatory power (Adorno et al., 1950, p. 260) were selected from a pool of F Scale items on which there was unanimous agreement by the members of an advanced seminar class that they possess transcultural meaning and do not refer to historical events or situations with a special meaning either in the United States or in Israel. Each item had the following response categories: strongly agree, agree, agree slightly, disagree slightly, disagree, strongly disagree, with a possible score from 1 to 6.

Authoritarian personality questionnaire. The Gough *Pr* scale, consisting of MMPI items that differentiated significantly between prejudiced and nonprejudiced individuals and correlated significantly with the F Scale, was used as an index of authoritarian personality characteristics (Gough, 1951a, 1951b). An analysis of the items of this scale reveals that they cover the various needs and adjustment mechanisms mentioned by Adorno et al. (1950) as typical of the prejudiced individual (Gough, 1951b).

Prejudice scales. The subjects' ethnic prejudice was determined by means of two especially constructed Likert-type scales: a 19-item scale of prejudice against Moroccan immigrants (NP scale), and a 17-item scale of prejudice against the local Arab population (AP scale). The original version of each scale consisted of especially constructed items, similar in content to the items used by Adorno et al. (1950) for their Ethnocentrism and Negro prejudice scales. The final version of each scale consisted of all items that correlated (r) .65 or higher with the subjects' total score on the original version of each scale. Test-retest reliability (r) was .91 and .89, respectively, for the two scales. The scoring system was the same as for the F Scale.

RESULTS

Table 1 gives means and standard deviations for the various scales and questionnaires. There was no significant sex difference on any of the measures. (The t values ranged from .13 to .38.)

Table 2 presents the intercorrelations (r) between the various measures. A highly significant correlation was obtained between the two prejudice scales. Only one of the prejudice scales, the AP scale (Arab prejudice) was significantly related to the F Scale. Both prejudice scales were significantly related with the Gough *Pr* scale, measuring

[1] This study was aided by a grant from the Committee for the Encouragement of Research, the Israeli Labor Organization.

The author is grateful to Jacob Jonah and Tsadok Baruch for their help in the collection of the data of this paper.

[2] Now at the University of Maryland School of Medicine.

[3] The following items were used: 1, 8, 13, 18, 21, 26, 27, 31, 37, 42 (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950, p. 260).

TABLE 1
SCALE STATISTICS

Measure	*M*	*SD*
F scale	39.98	9.67
NP scale	45.55	14.44
AP scale	57.89	18.35
Pr scale	10.37	4.38

TABLE 2
INTERCORRELATIONS BETWEEN THE VARIOUS MEASURES

	AP scale	F Scale	*Pr* scale
NP scale	.66**	.23	.39**
AP scale		.44**	.53**
F Scale			.53**

** Significant at .01 level.

authoritarian personality traits. A highly significant correlation was obtained between the F Scale and the Gough *Pr* scale.

DISCUSSION

These results provide cross-cultural validation for the major conclusions of *The Authoritarian Personality*. The high correlation between the subjects' Arab and North African prejudice scores is consistent with the hypothesis that ethnocentrism is a generalized trait. The fact that the Gough *Pr* scale, a measure of authoritarian personality needs and defenses, was a significant source of variance in both types of prejudice is certainly consistent with the basic theme of *The Authoritarian Personality*. The results of the present study do suggest, however, that not all ethnic prejudice is related to authoritarian ideology. The failure to obtain a significant correlation between the F Scale and North African prejudice, may be attributable to the independence of the latter from political affiliation.

The significant positive correlation between the F Scale and the Gough *Pr* scale is consistent with the hypothesis that personality needs and defenses determine attitudes, although the correlation cannot prove the supposed causal relation. An alternate possibility is that children learn an authoritarian ideology from parents who share it, and who, in keeping with their ideology, engage in socialization procedures that produce essentially authoritarian personalities. That is, rather than reflecting personality needs and defenses, ideology may influence socialization practices which in turn influence personality needs and defenses. The fact, however, that in the present study anti-Moroccan prejudice was related to the personality characteristics indexed by the Gough scale without being part of an authoritarian ideology as measured by the F Scale does suggest the former causal sequence.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.

CHRISTIE, R. Authoritarianism re-examined. In R. Christie & Marie Jahoda (Eds.), *Studies in the scope and method of "The authoritarian personality."* Glencoe, Ill.: Free Press, 1954. Pp. 123–196.

GOUGH, H. G. Studies of social intolerance: II. A personality scale for anti-Semitism. *J. soc. Psychol.*, 1951, **33**, 247–255. (a)

GOUGH, H. G. Studies of social intolerance: III. Relationship of the *Pr* scale to other variables. *J. soc. Psychol.*, 1951, **33**, 257–262. (b)

HYMAN, H. H., & SHEATSLEY, P. B. "The authoritarian personality"—A methodological critique. In R. Christie & Marie Jahoda (Eds.), *Studies in the scope and method of "The authoritarian personality."* Glencoe, Ill.: Free Press, 1954. Pp. 50–122.

(Received December 5, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 63, No. 3, 656-659

EFFECTS OF FAMILIARITY WITH A CONTROVERSIAL ISSUE ON ACCEPTANCE OF SUCCESSIVE PERSUASIVE COMMUNICATIONS

EDWIN J. THOMAS, SUSAN WEBB, AND JEAN TWEEDIE[1]
University of Michigan

ONE of the significant conclusions of *The Order of Presentation in Persuasion* by Hovland, Mandell, Campbell, Brock, Luchins, Cohen, McGuire, Janis, Feierabend, and Anderson (1957) denies the existence of a general law of primacy in persuasion—of any universal tendency for the first of two successive, persuasive communications to be more effective than the second. This conclusion, however, is based largely upon experiments in which familiar topics were used (e.g., protective tariffs, reduction of the legal voting age, socialized medicine), and familiarity with controversial topics is most generally accompanied by previous exposure to persuasive communication concerning the issues. Clearly it is not possible to study the effects of the *first* exposure to a persuasive message if a person is already familiar with the topic and its controversial features, although the immediate serial effects of two-sided persuasion can be studied with familiar topics. Thus the question of whether there is a general law of primacy in persuasion must be examined in experiments making use of unfamiliar topics.

With a new topic, the first persuasion has the possible advantages over the second of being more fully attended to, of being learned with less interference, of being accepted more readily, and of being established as a first impression that structures the perception of the subsequent persuasive message. The operation of any or all of these could generate primacy. Particularly relevant to this problem are studies by Luchins (Hovland et al., 1957), who attributed to set or Einstellung the marked primacy effect found in forming conceptions of an imaginary person from a limited description of him. While Luchins' studies had the virtue of dealing with the unfamiliar topic of a previously unknown person, they did not involve a controversial issue. The experiments reported here compare the effectiveness of successive pro and con persuasions with unfamiliar and familiar controversial topics to test the hypothesis that primacy should occur with an unfamiliar topic and that there should be no primacy or recency with a familiar topic.

[1] The authors wish to express their gratitude to Lynette Beall, Sharon Edwards, Cheryl Howard, Louis Jaffe, Daniel Marcus, Gail Smith, James Schreuder, Fred Steel, Terry Sokay, and Elizabeth Sweebe for help provided in various phases of the experiments.

EXPERIMENT I

Method

The subjects were 38 male and 24 female, high school seniors who participated in the experiment as "captive volunteers" while attending one of their classes.

Topic. The issue was whether to use Dr. Murray's new anti-cancer serum routinely in medical practice. To familiarize the subjects with it, they read a report on Dr. Murray's research essentially as it had appeared a few months earlier in an issue of a popular weekly news magazine; the brief described general facts concerning cancer, cancer research, how Dr. Murray developed the serum from horse blood, and some elements of the controversy concerning the wisdom of using this serum in medical practice. In the nonfamiliar condition the filler article was one on the discovery of a new tranquilizing tea, taken from the same magazine.

The arguments were not labeled as controversial; the pro position was entitled "An Anti-Cancer Serum" in which it was advocated that Dr. Murray's anti-cancer serum be used routinely in medical practice, and the con arguments were entitled "Information On the Use of Dr. Murray's Anti-Cancer Serum" and advocated withholding the use of serum from use in medical practice until further inquiry into its effects had been undertaken.

Response measures. The three measures of attitude regarding the use of the serum routinely in medical practice were nine-point Likert-type scales ranging from strongly disagree at 1 to strongly agree at 9, with 5 being neutral. The crucial question was "An anti-cancer serum recently developed by Dr. Murray of Canada should be routinely used in medical practice" and was embedded among six other questions, some of which were unrelated to the serum.

A multiple-choice knowledge test composed of 14 items was given to determine the effectiveness of the familiarity induction. The knowledge score consisted of the number of correctly endorsed items of the 10 dealing with significant facts of the familiarity induction; the remaining 4 items related to the tranquilizing tea.

Procedure. There were 17 subjects who received the pro-con order and 14 the con-pro in the familiar condition, and 16 who received the pro-con order and 15 the con-pro order in the nonfamiliar condition; males and females were represented approximately equally in each.

Subjects took one hour to complete the following tasks in a booklet: Attitude Questionnaire 1, familiarity induction (article on Dr. Murray's serum or filler article), Attitude Questionnaire 2, knowledge test, persuasive communications, and Attitude Question-

naire 3. While sitting one seat apart in the classroom, they went through the booklet without knowing what was to come subsequently, and after each page of responses was completed, the subject was directed to raise his hand so that it could be picked up.

Results

Success of inductions. The score on the knowledge test was 5.65 in the familiar group and 1.6 in the nonfamiliar group ($p < .01$ by t test), indicating that subjects in the familiar group generally became familiar with the topic and that subjects in the nonfamiliar group were previously unfamiliar with the topic. The familiarity induction also changed attitudes in a con direction; in the familiar condition the means on Attitude Questionnaire 1 and 2 were 5.06 and 4.39, respectively ($t = 59.5$, $p < .05$ by Wilcoxon's test); the analogous means for the nonfamiliar condition were 4.61 and 4.55, respectively. However, the scepticism regarding the use of the serum caused by the familiar induction cannot be regarded as disposing subjects in that condition toward either a primacy or recency effect.

The persuasive communications succeeded in moving the majority of the subjects from their original opinions. In the familiar condition, 68% of the subjects changed their opinion on Attitude Questionnaire 2, 71% changed from Attitude Questionnaire 2 to 3, and 77% had different positions on Questionnaire 3 as compared to Questionnaire 1; analogous percentages for the nonfamiliar condition were 45, 87, and 87, respectively. The mean amount of movement fell between one and three scale steps, with the nonfamiliar condition bringing about somewhat more movement than the familiar condition (p is not significant).

Attitude change. The primacy score computed for each subject was the amount of change, in the direction of the position advocated first, from the score obtained before the persuasive communications to the score gotten afterwards. When these scores were ranked over both conditions and the familiarity and nonfamiliarity groups compared by the Mann-Whitney U test, there was no difference found for changes from Attitude Questionnaire 1 to 3 or from Attitude Questionnaire 2 to 3.

EXPERIMENT II

One reason why the expected primacy effect did not appear in the nonfamiliarity condition in the first experiment is that the measurement of the effects of the persuasive communications occurred too soon after the second persuasive communication. Miller and Campbell (1959) have shown that if a general primacy effect exists it is most likely to be manifested with delayed rather than immediate measurement. The Miller-Campbell theory assumes that a persuasive communication may be forgotten like anything else and that persuasive messages are forgotten much more rapidly immediately after their initial learning than at later periods. Like the Ebbinghaus forgetting curves that served as models for the analogy, these forgetting curves take the form of a negatively accelerated function with time. Given this shape of the forgetting curve and the additional assumption that the strength of the first persuasive communication is stronger than the second (thereby elevating the curve for the first message), then the early measurement of the effects of opposing, persuasive communications may obscure the effects of the first message because of the relatively high asymptote of the more recent second message. Later measurement of these effects would reveal primacy because the asymptote for the first message would have settled at a higher level than that for the second message (Miller & Campbell, 1957, p. 2, Figure 2). To examine the possibility that a primacy effect was obscured in the nonfamiliar condition by too early measurement a second experiment was done replicating the first and adding delayed measurement of the effects of the persuasion communications.

Method

Subjects were 75 female undergraduates at the University of Michigan who participated in the experiment as part of a requirement for an introductory course in psychology. In the condition of immediate-measurement there were 17 subjects in the nonfamiliar group and 19 in the familiar group; in the condition of delayed measurement there were 17 subjects in the nonfamiliar group and 22 in the familiar group. The topic, procedure, and response measures were the same as in Experiment I, with the following exceptions: Attitude Questionnaire 2, given in the first experiment following the familiarity induction, was eliminated to reduce the number of times the subject had to give his attitude, thereby hopefully reducing any possible consistency effect. Two additional attitudinal items were added to the questionnaires given to the subjects before and after the persuasive communications, and five additional items were inserted into the final questionnaire. All of the items were nine-point Likert scales having the same form as described for those used in Experiment I. A question was inserted concerning the extent to which subjects perceived the pro and con arguments as controversial. Luchins' work (Hovland et al., 1957) suggests that subjects who perceive the statements as noncontroversial would be more likely to view the communications as a single unit than subjects who view the arguments as controversial. Communications seen as comprising a noncontradictory whole should be more subject to Einstellung structuring, and hence a primacy effect, than communications viewed as separate arguments.

Results

Success of inductions. The mean score on the knowledge test was 6.75 in the familiar condition

and .54 in the nonfamiliar condition ($p < .0001$ by t test), again indicating the effectiveness of the manipulation of familiarity. The persuasive communications also were again effective in moving a majority of the subjects from their original position; percentages of subjects changing ranged from 51 to 83.

Attitude change. Using the same method to test the significance of the primacy scores as in Experiment I, no differences were found on any of the three items given before and after the persuasive messages for the familiar and nonfamiliar groups in the immediate condition, the familiar and nonfamiliar groups in the delayed condition, and the nonfamiliar groups in the immediate and delayed conditions; the five questions used in the post questionnaire only were also analyzed, and again no differences were found.

Effect of perceived controversy. Exploration of Luchins' Einstellung interpretation of primacy was undertaken by comparing subjects who perceived the two sides of the argument concerning Dr. Murray's serum to be controversial versus those who did not. If Luchins' hypothesis is correct, then those subjects who did not perceive controversy between the pro and con arguments would be most likely to evidence a primacy effect and those who perceived that the issues were controversial would be least likely to evidence primacy. Fifty-three of the subjects saw the arguments as conflicting and 23 did not; there were no differences in primacy scores between these groups on any of the attitude items for either delayed or immediate measurement. We also examined the possibility that the primacy effect predicted for the nonfamiliar condition had occurred for subjects not perceiving controversy, but had not occurred for subjects perceiving controversy. The analysis revealed no unequivocal primacy effect for subjects not perceiving controversy in the nonfamiliar condition.

EXPERIMENT III

The condition of immediate measurement of the second experiment was replicated with 216 female undergraduates of the University of Michigan, with negative results.

DISCUSSION

Suggestions concerning how to interpret the failure to find a primacy effect in our studies can be derived from comparing our results with those of Miller and Campbell (1959), whose experiment appeared in print after the first of our experiments was conducted. The results of their experiment directly comparable to ours involve the comparison of the attitude-change scores for the conditions of immediate and delayed measurement. The topic was the question of who to favor in a suit for damages resulting from a defective vaporizer—an unfamiliar topic. They found no tendency toward primacy or recency in the condition of immediate measurement but a primacy effect for the condition of delayed measurement. Thus this study provides results corroborating the hypothesis of the present studies.

A critical difference between our studies and theirs was the topic. Although both topics were unfamiliar to the subjects, ours could readily have aroused existing attitudes toward the treatment of cancer, the chronically ill, taking risks in medical practice, and the related topic of euthanasia, whereas theirs involved a neutral and unknown subject. Thus the failure to obtain a difference in the condition of delayed measurement of our second experiment may have been due to the relative weakness of the persuasive communications in the face of the existing attitudes. This interpretation suggests the necessity to distinguish between familiarity with the topic and familiarity with the attitudes that may apply to it. The topic *as well as* attitudes relating to it should be new for the subjects, as presumably was the case in the Miller-Campbell experiment. If there is a new topic but it is similar to others known to the subjects and arouses attitudes that subjects already have toward the related topics, then the reactions of the subjects to the new topic should be a function of the extent to which it is similar to the already known topics and of the intensity of the attitudinal affect associated with the related attitudes.

SUMMARY

The hypothesis of these experiments was that the first of two successive persuasive communications would be more effective than the second, where subjects were not familiar with the topic, and that no primacy or recency effect would be found where subjects were familiar with the topic. In the first experiment, in which conditions of familiarity and nonfamiliarity with a controversial topic were established, no tendency was found toward either a primacy or a recency effect in either condition. In the second experiment the conditions of the first were replicated with the addition of delayed measurement added to that of immediate measurement. Again no tendency toward primacy or recency was found. A third experiment replicated the condition of immediate measurement of the second, again finding no primacy effect. Although the topic in these experiments was not familiar to the subjects, strong

existing attitudes toward related topics may have been aroused that made a primacy effect less likely.

REFERENCES

HOVLAND, C. I., MANDELL, W., CAMPBELL, ENID H., BROCK, T., LUCHINS, A. S., COHEN, A. R., MCGUIRE, W. J., JANIS, I. L., FEIERABEND, ROSALIND L., & ANDERSON, N. H. *The order of presentation in persuasion*. New Haven: Yale Univer. Press, 1957.

MILLER, N., & CAMPBELL, D. T. Recency and primacy in persuasion as a function of the timing of speeches and measurements. *J. abnorm. soc. Psychol.*, 1959, **59**, 1–10.

(Received December 5, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 63, No. 3, 660-662

ANXIETY AND THE EXPERIMENTAL AROUSAL OF AFFILIATION NEED[1]

DONN BYRNE[2]
University of Texas

THE effect of experimentally aroused affiliative motivation on responses to projective tests has been investigated in order to develop a TAT scoring system for n Affiliation. Shipley and Veroff (1952) employed a sociometric procedure as the arousing condition. Their scoring system focused on imagery involving affiliation-deprivation. They cross-validated their scoring categories by comparing responses of freshmen accepted and rejected by college fraternities. Atkinson, Heyns, and Veroff (1954) attempted to broaden the definition of affiliation imagery by including any statement of concern over establishing, maintaining, or restoring a positive affective relationship with another person or persons. They also maximized the difference between the experimental and control groups used in the procedure for analyzing the responses. Again, a sociometric rating experience constituted the arousal condition. In both studies, and with somewhat different scoring systems, more affiliation responses occurred in the stories of the aroused subjects.

The experimental conditions that have defined affiliation arousal have been such that both positive and negative cues probably are present. Affiliation need is conceptualized as an approach tendency, and the score as it is defined by the Michigan group is largely oriented toward approach responses. However, imagery motivated primarily by fear of rejection may have been confounded with that motivated primarily by hope of affiliation because of the nature of the arousal conditions that have been used. If arousal conditions that involve ratings by others actually do evoke avoidance responses, such conditions should evoke feelings of anxiety as well as affiliation need.

A further consideration is whether affiliation arousal would be expected to evoke anxiety to the same extent in all subjects. When TAT protocols are obtained in a neutral situation and scored for n Affiliation, the resulting individual differences are assumed to indicate variations in characteristic motivational states. Persons who are found to have high need for affiliation should be most responsive to affiliation arousing cues in their environments. That is, external stimulus conditions and internal motivational state should interact. French and Chadwick (1956), using their own projective measure of affiliation need, found the greatest increase in n Affiliation in an arousal condition for subjects whose scores were highest in a neutral condition. Perhaps experimental manipulations designed to arouse affiliative motivation are only effective with those subjects for whom such environmental cues are initially salient. Therefore, if anxiety is evoked by affiliation arousal, one might reason that its intensity should be a function of an individual's characteristic level of n Affiliation.

These two considerations lead to the hypotheses that (*a*) anxiety is evoked by the experimental arousal of affiliation motives and (*b*) the amount of anxiety is a positive function of n Affiliation determined in a neutral situation.

METHOD

Procedure. As part of another study, 84 students (44 males, 40 females) enrolled in introductory psychology courses at the University of Texas were seen in groups of 5–10. They were seated in a small amphitheater equipped with a large one-way mirror. They were given a paper-and-pencil task requiring about 15 minutes time.

For the experimental group, the following instructions were given:

> Before you begin, there is something else I would like to tell you about. You are actually taking part in two experiments at the same time.
>
> One way to make judgments about people is to observe them carefully. The study of expressive movements has taught psychologists a lot about the meaning of the many little unconscious gestures, facial expressions, body movements, etc. in which all of us engage in our everyday behavior. We have some evidence that such things have a lot to do with a person's popularity. So, during the rest of the session today, I and my two assistants will be watching each of you very closely.
>
> This mirror here is actually a one-way screen. In the room next to this one, two research assistants are already taking notes on various aspects of your behavior. They are both graduate students in psychology. I'll show you where they are. (Lights in experimental room are turned off and those in the observation room turned on momentarily.)
>
> Before the hour is over, we will have assigned a rating to each of you with respect to your general popularity, your attractiveness to the opposite sex, how likeable you seem, etc.

[1] This investigation was supported in part by research grants (EF-140, EF-143) from the University of Texas Research Institute.

[2] The author would like to express his appreciation to John Sheffield and Carole Calvert who served as research assistants and to Elaine Abbott who scored the affiliation need protocols.

Following this, the experimenter left the subjects alone and noisily entered the observation room.

Anxiety ratings. Immediately following the paper-and-pencil tasks, the experimenter returned and administered a questionnaire dealing with reactions to the experiment. Two six-point self-rating scales dealt with their feelings of nervousness and uneasiness during the preceding 15 minutes and their estimates as to how nervous or uneasy most of the other students in the room probably felt. On each scale, the descriptive points were "completely calm," "relatively calm," "a little uneasy," "quite uneasy," "very uneasy," and "extremely uneasy."

The anxiety score consists of the combined responses to these two scales. The split-half reliability of this instrument was found to be .72, corrected by the Brown-Spearman formula.

Affiliation need. One week later, a different experimenter using a different experimental room administered four TAT slides according to Atkinson's (1958, pp. 836–837) instructions. The slides consisted of Pictures 28, 83, 102, and 103 in the list presented by Atkinson (1958, pp. 832–834). A research assistant used the Heyns, Veroff, and Atkinson (1958) manual to score the protocols.

A subsample of 29 protocols was scored independently by an experienced scorer at the University of Michigan. The correlation of .95 between the two sets of scores indicates high interjudge reliability. For the 84 subjects, the mean n Affiliation score was 2.67 with a standard deviation of 2.54. Subjects were divided approximately at the median into low (N = 38, scores of 0-2) and high (N = 46, scores of 3-10) groups.

RESULTS

The means and standard deviations of the anxiety ratings are presented in Table 1. The interaction between experimental conditions and affiliation need in evoking anxiety was examined in a two-way analysis of variance. Disproportionality among the number of cases in the four cells was corrected using the procedure outlined by Wert, Neidt, and Ahmann (1954). The results of this analysis are reported in Table 2.

Neither n Affiliation nor the experimental conditions were related to self-ratings of anxiety.[3] The interaction between the two independent variables, however, was statistically significant. The major source of the difference is between the High and Low n Affiliation groups in the experimental condition. Subjects high in affiliation need indicated significantly greater anxiety in response to affiliation arousing cues than subjects low in affiliation need (t = 2.61, df = 34, $p < .02$). Also, the difference between the experimental and control conditions approaches significance for the subjects high in n Affiliation (t = 1.90, df = 44, p between .05 and .10).

[3] It was the failure of the arousal instructions to evoke anxiety that led to the present investigation.

TABLE 1
MEANS AND STANDARD DEVIATIONS OF ANXIETY RATINGS

	Control group			Experimental group		
	N	*M*	*SD*	*N*	*M*	*SD*
High n Affiliation	27	3.81	1.16	19	4.42	.88
Low n Affiliation	21	4.00	1.23	17	3.59	.97

TABLE 2
SUMMARY OF ANALYSIS OF VARIANCE OF ANXIETY RATINGS AS INFLUENCED BY EXPERIMENTAL CONDITIONS AND AFFILIATION NEED

Source	*SS*	*df*	*MS*	*F*	*p*
Experimental conditions	.4072[a]	1	.4072	.33	*ns*
Affiliation need	1.3438[a]	1	1.3438	1.09	*ns*
Interaction	5.2843[a]	1	5.2843	4.28	$<.05$
Within (error)	98.8233	80	1.2353		

[a] Adjusted.

Another way of examining the data is through a correlational approach. Anxiety ratings and n Affiliation scores correlated .45 ($p < .01$) in the experimental condition and $-.12$ (*ns*) in the control condition.

DISCUSSION

The results suggest that manipulations designed to arouse affiliation need evoke anxiety, at least in persons whose fantasy productions give evidence of concern over establishing, maintaining, or restoring positive affective relationships. Arousal conditions that involve a high probability of success should be less likely to evoke such anxiety than arousal conditions that involve the strong possibility of failure. A separation of these positive and negative qualities in the arousal condition might lead to a more specific differentiation of hope of success and fear of failure with respect to affiliation imagery in thematic material. As Shipley and Veroff (1952) have suggested, the ability to specify whether individuals are motivated primarily by positive, negative, or mixed affiliative needs would seem to permit more accurate behavior predictions in affiliation relevant situations.

From another viewpoint, the findings suggest that the present n Affiliation scoring system identifies those individuals who are likely to be made anxious by a situation containing an affiliation threat. Presumably these same high n Affiliation subjects would respond with greater pleasure to the promise of an affiliative reward than would subjects

low in n Affiliation. Other motivational scoring schemes might also be used in a similar way to predict those who would be made anxious by threat relevant experimental conditions or made pleased by reward relevant conditions.

A word should be added about the failure of the experimental condition to evoke more anxiety than the control condition. Remarks by various subjects suggested that the room in which the experiment was conducted was somewhat anxiety provoking in and of itself. If so, perhaps the instructions only served to define the source of anxiety cues. Without those instructions (in the control group), the array of individual hypotheses about the threat presented by the room was unrelated to n Affiliation. However, when other subjects (in the experimental group) were informed that they were being observed and rated, the source of concern became relatively specific. Those who feared, for example, the possibility of electric shock were now able to relax while those with strong affiliation motives perceived the cues as increasingly stressful.

SUMMARY

The experimental conditions employed in studies of n Affiliation probably contain a mixture of positive and negative cues. If so, such arousal conditions should evoke anxiety as well as n Affiliation. The hypothesis was tested that anxiety is evoked by the experimental arousal of affiliation motives and that the amount of anxiety is a positive function of n Affiliation determined in a neutral situation.

In the experimental group, subjects were aware of being watched through a one-way mirror and rated for popularity, attractiveness, etc. while completing a paper-and-pencil task. The control group simply completed the paper-and-pencil task. All subjects then rated themselves and others in the group with respect to feelings of nervousness and uneasiness. At another testing session, TAT slides were shown, and the protocols were scored for n Affiliation.

Experimental conditions and affiliation need interacted to influence the anxiety self-ratings. In the experimental group, subjects high in n Affiliation rated themselves significantly more anxious than did those subjects low in n Affiliation; in the control group, affiliation need was unrelated to anxiety ratings. Manipulations designed to arouse affiliation need thus appear to evoke anxiety in persons whose fantasy productions give evidence of concern over affiliation. Also, the present n Affiliation scoring system identifies those individuals who are made anxious by an affiliation threat.

REFERENCES

ATKINSON, J. W. (Ed.) *Motives in fantasy, action, and society.* Princeton, N. J.: Van Nostrand, 1958.

ATKINSON, J. W., HEYNS, R. W., & VEROFF, J. The effect of experimental arousal of the affiliation motive on thematic apperception. *J. abnorm. soc. Psychol.*, 1954, **49**, 405–410.

FRENCH, ELIZABETH G., & CHADWICK, IRENE. Some characteristics of affiliation motivation. *J. abnorm. soc. Psychol.*, 1956, **52**, 296–300.

HEYNS, R. W., VEROFF, J., & ATKINSON, J. W. A scoring manual for the affiliation motive. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society.* Princeton, N. J.: Van Nostrand, 1958. Pp. 205–218.

SHIPLEY, T. E., & VEROFF, J. A projective measure of need for affiliation. *J. exp. Psychol.*, 1952, **43**, 349–356.

WERT, J. E., NEIDT, C. O., & AHMANN, J. S. *Statistical methods in educational and psychological research.* New York: Appleton-Century-Crofts, 1954.

(Received December 9, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 63, No. 3, 663-665

PSYCHOGENIC DEAFNESS, PERCEPTUAL DEFENSE, AND PERSONALITY VARIABLES IN CHILDREN

G. F. REED

University of Manchester

In a recent discussion, Kodman and Blanton (1960) have suggested examination of the perceptual defense hypothesis, using as subjects children who manifest psychogenic auditory behavior. Indicating the theoretical difficulties involved, they observe that a preponderance of the cases encountered by them were children whose normal thresholds for speech contrasted with losses on the pure tone audiogram, i.e., in response to *meaningless* stimuli.

It may be postulated that conventional pure-tone audiometry presupposes conditions of set and attention different in quality from those obtaining in normal speech communication. If this is so, then personality may well be an important variable in the determination of responses in the two situations. The present study reports a preliminary check on this prediction, with some observations on the applicability of perceptual defense thories to psychogenic deafness.

Procedure

The files (totaling over 3,000) of a large city child guidance clinic and school psychological service were searched for cases presenting psychogenic or functional deafness for speech where responses to pure-tone audiometric testing had been judged to be normal. A further group of children was found whose thresholds for pure tones had been relatively high but whose responses to speech had been within normal limits. In both groups the age range was restricted to 8-14, and cases of below 75 IQ, those where there was any evidence of brain injury, and those where the psychiatric diagnosis was in doubt were excluded. All cases were psychiatrically diagnosed, independently of this investigation, as being predominantly "hysteric" or "anxiety state." The children in the first group presented outgoing behavior problems, whilst those in the second were characterized by timidity, withdrawal, or mild phobic symptoms.

Both groups were roughly equal in distributions of age, sex, and IQ.

Results and Discussion

Findings are summarized in Table 1. Despite the statistically high significance ($p = .005$), confidence in generalization is, of course, diminished by considerations of the small numbers involved and the unreliability of psychiatric diagnosis. But there seems to be a clear tendency in this sample for those children whose auditory threshold for pure tones is high to be categorized as hysteric (as opposed to anxious) and some tendency towards the reverse among those who display a lack of response to speech.

Kleinman (1957), in a careful experimental comparison of subjects with psychogenic hearing loss and a control group with organic hearing loss, found a decreased practice effect rather than heightened auditory thresholds in the experimental group. However, Kleinman pointed out that "individuals whose auditory thresholds are characteristically raised as a defense against anxiety-provoking stimuli can be identified on the basis of their perceptual behaviour," and interpreted his results as being generally consistent with perceptual defense theory. Other workers have concurred with this view, and it may well be presumed that some relationship does exist between psychogenic deafness and perceptual defense. Indeed, investigations into the latter have, to some extent, been stimulated by clinical findings. But theoretical difficulties at once become apparent, as Kodman and Blanton (1960) have pointed out.

Most relevant experimental studies have been designed to show whether there are differences between subjects' responses to "threatening," as opposed to neutral, *words or sentences*. In practice, our cases reacted abnormally to *situations*. Group A showed high or inconsistent thresholds for pure tones during audiometric testing, but not for speech, Group B members were hard of hearing among strangers or in class, but responded normally to pure-tone testing. Thus it is difficult to explain the behavior of Group A in terms of perceptual defense as presented in Eriksen's (1954) review or in Blum's (1955) more specifically psychoanalytic account unless it is postulated that pure-tone audiometry constitutes a threatening stimulus.

On the other hand, Solomon and Howes' (1951) argument regarding the relative lowness of appropriate responses in the subjects' response hierarchies might well explain the high thresholds of Group A, but not those of Group B. In what to them are presumably anxiety provoking situations, these children show a relative lack of response, not only to unfamiliar or "tabu" words, but to the bulk of verbal stimuli.

In short, the concept of perceptual defense in its original forms is not directly applicable to the behavior of children in Group A. Their type of response is more readily studied in terms of labora-

TABLE 1
DISTRIBUTION OF CHILDREN WITH PSYCHOGENIC DEAFNESS BY PERSONALITY TYPE

	Anxiety state	Hysteric	Totals
Elevated thresholds for:			
Pure Tones (Group A)	1	13	14
Speech (Group B)	9	3	12
	10	16	26

χ^2 (after Yates' Correction) = 9.86

tory and industrial studies of "vigilance" as defined by Mackworth (1957). Furthermore the early experiments on perceptual defense do not seem to have a direct bearing upon the behavior of children in Group B unless wide generalizations from the early theory are to be permitted. If the behavior of both groups is in fact explicable in terms of perceptual defense, then we would suggest, as did Singer (1956) in a lucid review and experimental study, that the most rewarding type of approach is one using "hypothesis" concepts as presented, for instance, by Postman (1953). Crudely put, it may be argued that some children adopt a generalized anxious set for certain social situations, and thus show elevated thresholds for all verbal stimuli in those situations. Other children adopt faulty expectations in perceptual tasks demanding heightened attention, and thus produce insensitive or inconsistent responses. It remains now to attempt to account for the individual differences in response to these two types of situation, which are highlighted by the distributions of clinical categories in the present data.

Where personality variables are concerned, the present findings, for reasons given above, cannot be directly compared with those of Lazarus, Eriksen, and Fonda (1951), who set out to study personality mechanisms as reflected in the "auditory recognition of sexual and aggressive material." It is of interest to note, however, that they found that their freely verbalizing or "intellectualizer" group, which included obsessive-compulsives, showed greater accuracy in the perception of threatening material than did their "repressor" group, which included conversion hysterics. Objection may be made to the clinical aptness of these workers' classificatory procedure. But their empirical results are clear and seem to point in the opposite direction to that indicated by the present data. Kurland (1954), on the other hand, found no difference between obsessional-compulsives and hysterics in their recognition threshold of emotional words. Again these results are not directly comparable with the present ones, but would appear to be different in implication.

The different distributions of clinical classifications in our two groups are predictable, however, from Eysenck's (1957) formulations. Pure-tone audiometry is a task which might be expected to discriminate between dysthymic individuals, with their relatively high levels of drive, aspiration, and persistence, and hysterics, who have been found to build up reactive inhibition more rapidly and to be behaviorally more distractible. In this connection, Kodman and Blanton (1960) point out the importance for accurate signal detection of freedom from competing response tendencies, and a most stimulating account of the differential performance of introverts and extraverts in relevant experimental tasks is given by Broadbent (1958).

A similar explanation, coupled with considerations of attention gaining behavior, might be given for the existence of a proportion of hysterics in Group B. As far as the majority of children in this group is concerned, however, a different set of mechanisms must be postulated; but, again, a plausible explanation is readily derived from Eysenckian theory. Dysthymic children are highly reactive and readily conditionable. They tend to demonstrate strong fear reactions, which, linked with oversocialization, may be triggered off by apparently neutral stimuli in social situations. Heightened thresholds for speech may thus be acquired as an anxiety reducing avoidance mechanism by these children more readily than by hysterics.

SUMMARY

Survey of a child guidance population showed that of 26 functionally hard-of-hearing cases, 14 presented elevated thresholds for pure tones but were within normal limits for speech. The other group, though normal on pure-tone testing, presented elevated thresholds for *all* speech, rather than for relatively threatening stimuli. It would appear that both groups were reactive to situations, as opposed to specific auditory stimuli within those situations. Furthermore the majority of the pure-tone group were clinically classified as "hysteric," whereas the majority of the speech group was classed as "anxiety state."

The applicability of "defense" theories to such cases is discussed, and it is suggested that the behavior of both groups may best be studied in the light of a "hypothesis" approach.

It is pointed out that the pure-tone testing situation demands "vigilance" in the experimental sense, and it is then shown that the differing

distributions of clinical categories in the two groups is predictable from Eysenck's formulations.

REFERENCES

BLUM, G. S. Perceptual defense revisited. *J. abnorm. soc. Psychol.*, 1955, **51**, 24–29.

BROADBENT, D. E. *Perception and communication*. London: Pergamon, 1958.

ERIKSEN, C. W. The case for perceptual defense. *Psychol. Rev.*, 1954, **61**, 175–182.

EYSENCK, H. J. *The dynamics of anxiety and hysteria*. London: Routledge & Kegan Paul, 1957.

KLEINMAN, M. L. Psychogenic deafness and perceptual defense. *J. abnorm. soc. Psychol.*, 1957, **54**, 335–338.

KODMAN, F., & BLANTON, R. L. Perceptual defense mechanisms and psychogenic deafness in children. *Percept. mot. Skills*, 1960, **10**, 211–214.

KURLAND, S. H. The lack of generality in defense mechanisms as indicated in auditory perception. *J. abnorm. soc. Psychol.*, 1954, **49**, 173–177.

LAZARUS, R. S., ERIKSEN, C. W., & FONDA, C. P. Personality dynamics and auditory perceptual recognition. *J. Pers.*, 1951, **19**, 471–482.

MACKWORTH, N. H. Some factors affecting vigilance. *Advanc. Sci.*, 1957, **53**, 389–393.

POSTMAN, L. On the problem of perceptual defense. *Psychol. Rev.*, 1953, **60**, 298–306.

SINGER, B. R. An experimental inquiry into the concept of perceptual defence. *Brit. J. Psychol.*, 1956, **47**, 298–311.

SOLOMON, R. L., & HOWES, D. H. Word frequency, personal values and visual duration thresholds. *Psychol. Rev.*, 1951, **58**, 256–270.

(Received December 20, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 63, No. 3, 666–670

AN OUTCOME INDEX FOR MENTAL HOSPITAL PATIENTS[1]

E. I. BURDOCK AND ANNE S. HARDESTY
New York State Department of Mental Hygiene and Columbia University

At any particular point in time, the patients in a mental hospital vary with respect to length of stay, number of admissions, number of releases, and time since first admission. A single index number combining these factors could provide a criterion for assessing the predictive validity of psychiatric prognoses, psychological tests, and treatment regimens.

Previous attempts to describe the fates of institutionalized patients by Crandell, Zubin, Mettler, and Logan (1954) and by Crandell, Zubin, Mettler, and Kugelmass (1956) culminated in an "immobility index," which was the ratio of number of months of hospitalization to number of admissions during the two years following first admission. This immobility index was found to correlate negatively with subsequent outcome.

The *Outcome Index*,

$$OI = \frac{t_o}{l} \times \frac{r^2}{a^5}$$

represents a refinement of the earlier immobility index in the direction of greater sensitivity and convenience. The four elements are:

- l = length of time in days between first admission and a selected anniversary date.
- t_o = total number of nights slept outside the hospital on official leave or discharge status.
- a = number of admissions.
- r = number of releases.

The relative weights given to admissions and releases were determined empirically from a sample of patients ranked in order of length and frequency of hospital experience. Previous studies as well as clinical observation suggest that each successive readmission increases the expectation of subsequent admissions, but each successive release is less indicative of genuine remission.

The Outcome Index is like a probability in that it ranges from zero for a patient with no significant releases to a hypothetical maximum of one for a patient who spent no days in the hospital.

The Outcome Index has been applied to three samples of patients first admitted to a New York State mental hospital in 1953, 1955, and 1956, respectively. They were followed up at four successive intervals: 6 months, 1 year, 2 years, and 3 years. The long-term follow-up was greatly facilitated by the fact that the patients had been located on the earlier occasions. The data were tabulated and Outcome Indices were computed for each patient at every period.

The 1953 sample consisted of 35 first-admission patients who were all young schizophrenics (mean age 29) and who had been selected by the professional staff for accessibility and amenability to psychological testing. The 1955 group of 60 first-admission patients was a sample of young schizophrenics (mean age 30) chosen by a survey of the case records. The 1956 group of 85 first-admission patients successively admitted to the same institution were heterogeneous from the standpoint of diagnostic category, age, sex, etc., being alike only with respect to time of entry into the hospital and first-admission status. These 85 patients were the survivors of a somewhat larger cohort reduced by deaths. (There appeared to have been a disproportionately large number of geriatric admissions in this cohort.) No Outcome Indices were computed for those who failed to survive to the end of the final follow-up period.

Previous studies (Kramer, Goldstein, Israel, & Johnson, 1955) have suggested that by the end of the second year of hospitalization, the careers of mental patients have become stabilized. Therefore, 2 years has been taken as a basal follow-up period for dividing patients into homogeneous groups with respect to the Outcome Index. Figures 1 through 3 represent the trends of individual Outcome Indices from 6 months to 3 years for each sample successively. In each sample, the patients have been divided into four groups according to the magnitude of the 2-year index. Chronic residents with zero indices at 720 days are not shown on the figures. There were 4 of these in the 1953 sample, 4 in the 1955 sample, and 17 in the 1956 sample. Only 3 of these 25 with zero index showed any change by the third year. One of the 1953 patients rose from zero to .004 and 2 of the 1956 patients rose from zero to .013 and to .222, respectively.

In each of the three figures, the right-hand graph, *Early Release*, represents patients with a 2-year Outcome Index of .50 or higher. To obtain a basal index this high, these patients must have had only a single admission and must have been released within one year. Some of these, with a 2-year index of .75 or higher, were released within 6 months of admission. In each of the three samples more than half the patients fit this

[1] This research has been supported by Research Grant M-3546 from the National Institute of Mental Health.

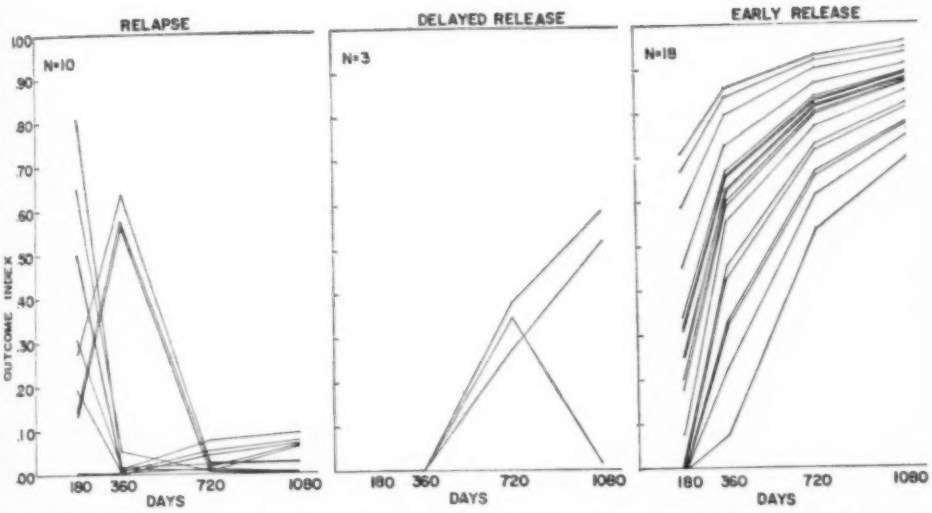

FIG. 1. Individual lines of trend of successive Outcome Indices from time of first admission to a state mental hospital in 1953.

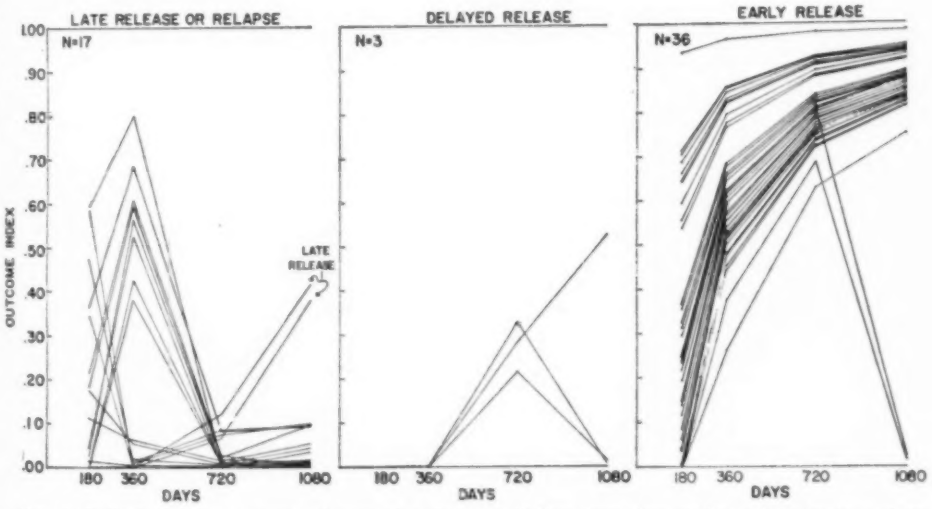

FIG. 2. Individual lines of trend of successive Outcome Indices from time of first admission to a state mental hospital in 1955.

definition for early release:

In 1953	18 out of 35
In 1955	36 out of 60
In 1956	44 out of 85
Total	98 out of 180

Only 4 of the 98 relapsed during the third year, 2 from the 1955 sample and 2 from the 1956 sample.

The middle graphs in Figures 1 through 3 are for patients with *Delayed Release*. The 2-year Outcome Index ranging from .125 to .50 again reflects a single admission, but with between 12 and 21 months of hospitalization. There were only three such patients in each sample.

The three left-hand graphs, *Late Release or Relapse*, are for patients with 2-year Outcome Index between zero and .125. A few patients with

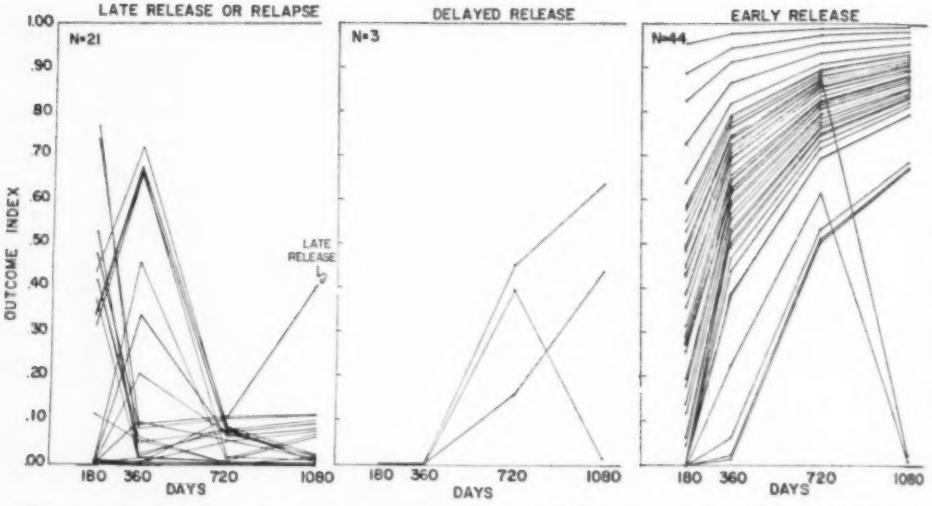

FIG. 3. Individual lines of trend of successive Outcome Indices from time of first admission to a state mental hospital in 1956.

Outcome Index less than .125 had a single admission but were released only after 21 months of hospitalization. There were two such patients in the 1955 sample and one in the 1956 sample. But the majority of those with an Outcome Index between zero and .125 were released during their first year and readmitted before the end of the 2-year basal period.

Figures 1–3 also suggest that the frequencies for each of the four categories of outcome are unequal. The disparity is more clearly revealed in Table 1, which presents the data in summarized form and includes the appropriate statistical tests of significance. The tests support the conclusion that in each sample the frequencies of the four different outcome categories are significantly different from each other. Moreover, an overall test of the pooled data, as shown in Table 1, indicates that the differences among the categories tend to be consistently in the same direction from sample to sample. Finally, a test of independence encourages the presumption that all three samples are drawn from the same population of outcomes.

The stability of the basal Outcome Index is most graphically demonstrated by dividing each sample into two subgroups, those with a 2-year index equal to or greater than .5 vs. those with a 2-year index less than .5. If separate frequency distributions of the 3-year Outcome Indices for each subgroup are then plotted on the same axes, the extent of overlap is seen to be negligible. Figures 4 through 6 show the data arranged in this way.

The evidence so far presented suggests that, in general, when it is applied to individual patients the basal Outcome Index computed for the 2-year period following first admission yields a maximum amount of information in a minimum follow-up time. However, it will often be of interest to examine the records of a cohort at the end of their first year so as to get some notion as to the

TABLE 1

CHI SQUARE ANALYSIS OF FREQUENCIES IN FOUR CATEGORIES OF A TWO-YEAR OUTCOME INDEX APPLIED TO THREE SAMPLES OF FIRST-ADMISSION PATIENTS AT A STATE MENTAL HOSPITAL

Sample	OI = 0	$0 < OI < .125$	$.125 \leq OI < .50$	$.50 \leq OI < 1.00$	Total	χ^2	*df*	*p*
1953	4	10	3	18	35	16.312	3	$<.001$
1955	4	17	3	36	60	47.332	3	$<.001$
1956	17	21	3	44	85	40.879	3	$<.001$
					$(\Sigma\chi^2)$	104.523	9	$<.001$
Pooled	25	48	9	98	180	100.310	3	$<.001$
				(Test of Independence)		4.213	6	*ns*

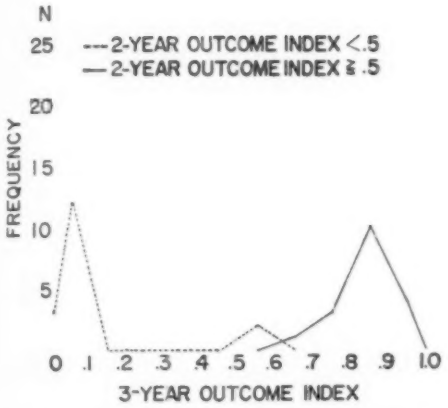

FIG. 4. Frequencies of various 3-year Outcome Indices among two subgroups of a 1953 sample of first-admission patients.

FIG. 5. Frequencies of various 3-year Outcome Indices among two subgroups of a 1955 sample of first-admission patients.

types of patients and their disposition. For this purpose the Outcome Index has been applied to a sample of 83 first-admission patients who entered New York Psychiatric Institute in 1958–1959. The New York Psychiatric Institute is a small intensive treatment facility which selects patients suitable for teaching and research and which does not relegate any patient to custodial care. It has a high staff-to-patient ratio. As it is still less than 2 years since admission, a basal Outcome Index could not be computed for these patients. But it is of interest to compare them with unselected state hospital patients for early outcome. The proportion of institute patients with zero index for the first 6 months is almost twice as large as the corresponding proportion for the state hospital; by the end of the year, the proportion of institute patients with zero index has increased so that the ratio is 5:2. The reason why fewer patients are released from the Psychiatric Institute during their first year appears to be that almost all patients become involved in intensive and prolonged psychotherapy (Burdock, Glass, Hardesty, & Beck, 1961).

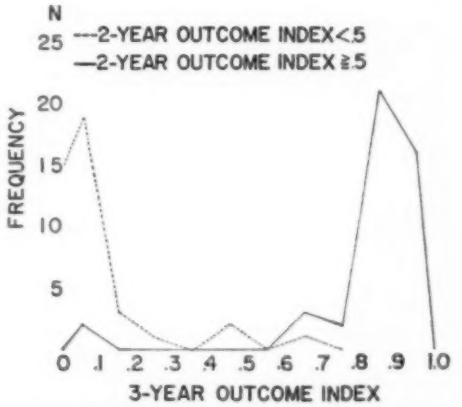

FIG. 6. Frequencies of various 3-year Outcome Indices among two subgroups of a 1956 sample of first-admission patients.

only one year of follow-up, the ranges of the three categories above zero are correspondingly reduced as follows:

OUTCOME INDEX	2-YR. FOLLOW-UP (BASAL PERIOD)	1-YR. FOLLOW-UP
OI = 0	Continuous residence	Continuous residence
0 < OI < .125	*Late release* (only after 21 months) or *Relapse* (2 or more admissions)	Release only after 10.5 months or Relapse
.125 ≦ OI < .50	*Delayed release* (only after 1 year)	Release after 6 months
.50 ≦ OI < 1.00	*Early release* (within first year)	Release within 6 months

The difference between the state hospital and the Psychiatric Institute is pointed up by tabulating the corresponding frequencies for each outcome category for the end of the first year after admission. When the Outcome Index is based on

Table 2 shows the relative frequencies of the 1-year outcome categories for each of the three samples from the state hospital and for the sample from the Psychiatric Institute. The accompanying statistical tests again support the hypotheses

TABLE 2

CHI SQUARE ANALYSIS OF FREQUENCIES IN FOUR CATEGORIES OF A ONE-YEAR OUTCOME INDEX APPLIED TO THREE SAMPLES OF FIRST-ADMISSION PATIENTS AT A STATE MENTAL HOSPITAL AND ONE SAMPLE OF FIRST-ADMISSION PATIENTS AT A PSYCHIATRIC INSTITUTE

Sample	OI = 0	0 < OI < .125	.125 ≤ OI < .50	.50 ≤ OI < 1.00	Total	χ^2	*df*	*p*
1953	7	8	5	15	35	6.485	3	*ns*
1955	9	7	8	36	60	39.337	3	<.001
1956	21	16	9	39	85	23.186	3	<.001
				($\Sigma\chi^2$)		69.008	9	<.001
Pooled	37	31	22	90	180	62.532	3	<.001
P.I.	41	8	12	22	83	31.361	3	<.001
				($\Sigma\chi^2$)		93.893	6	<.001
Pooled	78	39	34	112	263	61.029	3	<.001
Tests of Independence								
Among state hospital samples: $\Sigma\chi^2$ − Pooled χ^2 = 6.476							6	*ns*
State hospital vs. Psychiatric Institute: $\Sigma\chi^2$ − Pooled χ^2 = 32.863							3	<.001

that the four outcome categories are on the whole significantly different and that the variations are in the same direction from sample to sample, but the 1953 sample, the smallest, fails to provide a statistically significant difference. However, a noteworthy contrast emerges between the institute sample and the combined state hospital samples when the tests of independence are examined. While all three state hospital samples appear to come from the same population of outcomes, the Psychiatric Institute sample registers as in all probability independent. The institute's policy of holding patients for intensive psychotherapy is reflected in a slower release rate during the first year after admission.

From the data presented here it can be concluded that the Outcome Index, based on the first 2 years after admission, provides a reasonably stable criterion for assessing effectiveness of treatment and for distinguishing institutional policies. Three successive samples of first-admission patients taken from a regular state mental hospital show basically similar patterns of outcome despite different diagnoses, different age groups, and the varying effects of the current trend in drug therapy.

The Outcome Index can be used to characterize institutions as well as to describe individual patients. This latter application is illustrated by the comparison of the three state hospital samples with the sample from the Psychiatric Institute.

SUMMARY

An *Outcome Index* for mental hospital patients is described. The index relates time out of hospital to follow-up period since first admission and weights their ratio differentially by number of readmissions and releases. This index becomes relatively stable on 2-year follow-up. It can be used as a criterion for prognostic efficiency, for efficacy of treatment, and for characterizing institutional policy.

REFERENCES

BURDOCK, E. I., GLASS, LEAH, HARDESTY, ANNE S., & BECK, YVONNE. A sample survey of mental hospital patients. *J. clin. Psychol.*, 1961, **17**, 253–259.

CRANDELL, A., ZUBIN, J., METTLER, F. A., & KUGELMASS, N. "Mobility" in chronic schizophrenia with special regard to psychosurgery. *Psychiat. Quart.*, 1956, **30**, 96–113.

CRANDELL, A., ZUBIN, J., METTLER, F. A., & LOGAN, N. D. The prognostic value of "mobility" during the first two years of hospitalization for mental disorder. *Psychiat. Quart.*, 1954, **28**, 185–210.

KRAMER, M., GOLDSTEIN, H., ISRAEL, R. H., & JOHNSON, N. A. A historical study of the disposition of first admissions to a state mental hospital. Experience of the Warren State Hospital during the period 1916–1950. *Publ. hlth. Monogr.*, 1955, No. 32. (Publ. Hlth. Serv. Publ. No. 445. Publ. concurrently w. Publ. hlth. Rep., 1955, **70**, 1135–1136)

(Received December 27, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 63, No. 3, 671-674

STATUS AND THE JUDGMENT OF OCCUPATIONAL PRESTIGE[1]

SIDNEY I. PERLOE
Haverford College

When people are asked to rate a series of stimuli on an arbitrary category scale, their judgments are made in relation to one or more reference points or anchors. The rating assigned to a stimulus depends upon the direction and magnitude of the distance between the stimulus and the anchor (Volkmann, 1951). A shift in the location of the anchor toward the high end of the rating dimension leads to a displacement of the stimulus judgments toward the low end. A displacement of stimulus judgments in an upward direction is found when the anchor is shifted toward the low end of the judgment dimension.[2] These judgmental phenomena have been observed with both physical and verbal stimuli (Guilford, 1954; McGarvey, 1943).

There are a variety of ways in which the locations of anchors are determined. Sometimes subjects spontaneously choose the most extreme stimuli in a series as anchors because they are most easily discriminated from the other stimuli. At other times the experimenter provides special stimuli that subjects are instructed to use as anchors. In both of these cases the anchor is an external stimulus that is continuously or intermittently present in the judgment situation. There have, however, been studies which suggest that people may also use internal anchors to guide their judgments. Hunt (1941) found that asking subjects to imagine the most extreme stimulus they could think of, on a given dimension, had the same effect as employing an external anchor. Tresselt (1948) compared the weight judgments of weight lifters, watchmakers, and students. She found that the weight lifters initially judged the experimental weights as lighter than did either of the other groups. Presumably this difference was due to heavier internal anchors on the part of the weight lifters than on that of the watchmakers or students.

[1] Supported in part by NSF Grant G.3023.

The author wishes to express his appreciation to P. Billard, G. Brauner, H. Stick, T. Warne, and W. Welsh of Yale College, for their help in collecting part of the Yale data.

[2] A related theory proposed by Helson (1948) states that the rating assigned to a stimulus is determined by the distance between the stimulus and the adaptation level or neutral point of the series. Helson holds that anchors displace judgments only insofar as they change the adaptation level. In the present case, Volkmann's (1951) and Helson's theory make the same predictions.

Internal anchor effects have also been found by Hovland, Harvey, and Sherif (1957); Hovland and Sherif (1952); Kelley, Hovland, Schwartz, and Abelson (1955); Manis (1960); Segall (1959); Sherif and Hovland (1953); and Weiss (1959). In these studies, the subjects' opinions acted as internal anchors in their judgments of the position advocated by a series of opinion statements. Subjects with extremely favorable positions on an issue tended to displace their judgments of items that deviated from their views toward the unfavorable end of the judgment scale. Those with extremely unfavorable positions tended to displace their judgments of statements that deviated from their views toward the favorable end. Similar results have been obtained by Hinckley and Rethlingshafer (1951) in a study of the effects of the subjects' own heights on judgments of height; a series of heights were judged to be shorter by tall men than by short men.

The present paper is concerned with social status as an internal anchor in the judgment of occupational prestige. The hypothesis is that high status subjects should displace their judgments of medium and low prestige occupations toward the low end of the occupational prestige dimension.

Method

The data to be presented have been assembled from two sources. The first is a survey conducted by the National Opinion Research Center (1953). The survey sample covered a nationwide cross-section of Americans whose occupations ranged from laborer to professional man. The respondents were asked to give their own personal opinions of the general standing, i.e., prestige, of each of a series of 90 occupations. The ratings, made on a five-point scale, were converted into prestige scores by assigning a weight of .2 to the percentage of respondents rating a given occupation in the lowest category, a weight of .4 for the next lowest, .6 for the middle, .8 for the next highest, and 1.0 for the highest. The weighted percentages were then summed across the categories to yield the prestige score for the occupation in question. By this procedure the minimum score was 20, assigned to an occupation unanimously placed in the lowest rating category; the maximum was 100, assigned to an occupation unanimously placed in the highest rating category.

The data for high status subjects were taken from three unpublished studies performed by the author with Yale undergraduates in which incidental ratings of occupational prestige were collected. All three required subjects to rate the amount of prestige com-

TABLE 1
PRESTIGE SCORES

Nationwide decile	N[a]	Sample occupation[b]	Nationwide mean	Yale mean	Difference	p[c]
10	6	Diplomat	92.17	93.33	−1.16	ns
9	5	Judge	86.60	88.90	−2.30	ns
8	5	Psychologist	84.00	80.00	4.00	.05
7	4	Symphony musician	79.75	69.00	10.75	.03
6	5	Farm owner and operator	75.00	58.60	16.40	.004
5	8	Policeman	70.13	56.37	13.76	.001
4	5	Plumber	63.80	41.40	22.40	.004
3	7	Restaurant cook	56.71	41.89	14.82	.001
2	5	Farm hand	49.20	27.00	22.20	.004
1	6	Shoeshiner	38.83	32.17	6.66	ns

[a] *N* refers to the number of occupations used in each decile of the complete nationwide distribution.
[b] The sample for each decile lies as close as possible to the median of the Yale ratings of occupations in that decile.
[c] One-tailed values.

manded by each occupation in the population at large.[3] The first study included 14 subjects whose data were appropriate for the present problem, the second 51, and the third 41. In the first study, 31 occupations were the same as the ones in the nationwide sample; in the second study 44; and in the third 14. There was considerable overlap in the occupations used in each of the studies so that, taken as a whole, the Yale sample rated 56 of the 90 occupations used in the nationwide survey. The data from each of the Yale samples were converted into the same kind of prestige scores employed in the survey. Where the same occupation was used with more than one Yale sample, the mean of the several scores was used to arrive at an overall value for the Yale group.

The hypothesis that high status subjects would displace their ratings of moderate and low prestige occupations toward the low end of the prestige dimension was tested by comparing the prestige scores based on the judgments of the nationwide sample to ones based on estimates of these judgments by Yale students. The comparison was made for the 56 occupations which were rated by both groups. The prestige scores of the nationwide sample can be considered an objective measure of social reality, much in the same way as the readings on a balance are an objective measure of physical reality. The Yale students' judgments of the prestige commanded by a given occupation in the population at large may be thought of as equivalent to judgments of the weight of a lifted stimulus. In the case of the latter, accuracy is defined in terms of agreement between the judgments and the reading on the balance. In the former, accuracy is defined by the agreement between the judgments and the scores of the nationwide sample.

RESULTS

Ten occupational prestige levels were created by grouping the occupations falling in the same decile of the prestige score distribution as found with the nationwide sample (see Table 1 for a sample occupation from each decile). The mean prestige scores of the nationwide and of the Yale samples were computed for each decile. These means are presented in Table 1.

In general, the data support the hypothesis that high status subjects displace their judgments of middle and low prestige occupations toward the low end of the prestige dimension. With the exception of the two highest deciles, the Yale sample produced lower prestige scores than the nationwide sample. That is, the Yale students erred by displacing their judgments of the occupations in the bottom eight deciles toward the low end of the scale. They gave lower prestige scores to 43 of the 45 occupations located in this region. An analysis of variance[4] showed that the overall difference between the Yale and nationwide scores was significant ($F = 124.66$, $p < .005$, $df = 1/92$) and that the interaction between the status of the judges, i.e., Yale or nationwide, and the prestige of the occupations judged was also significant ($F = 8.34$, $p < .005$, $df = 9/92$).

In order to interpret the significant interaction, the difference between the scores assigned by the two samples to occupations in each decile was tested for significance using the Mann-Whitney *U* test. In Table 1 it can be seen that the differences between the samples with occupations from deciles 2 through 8 were significant. The lack of significance for the lowest decile may have been due to the unambiguous nature of the prestige of the lowest occupations, which may have prevented judgmental displacements from occurring. This suggestion is similar to one offered by Hinckley and Rethlingshafer (1951) as an explanation for

[3] The specific instructions were based on those used by McGarvey (1943, p. 28).

[4] The analysis was based on the technique for unequal frequencies presented by Walker and Lev (1953, p. 381). The error term was computed from the variance among the scores of the several occupations located in each of the 20 cells of the design.

their finding that tall and short men differ significantly in their judgments of moderate heights, but not of very tall or very short ones.

Yale students appeared to be able to make finer discriminations near the high prestige end of the scale than near the low end. The differences between the ratings assigned to occupations in adjacent deciles were significant at $p < .05$ or better, two-tailed test,[5] for the following pairs: 8 and 9, 7 and 8, 4 and 5, 2 and 3; the difference between deciles 6 and 7 approached significance ($p < .12$). This result is generally consistent with previous findings of greater discrimination among stimuli near anchors than among those farther away (Volkman, 1951). However, contrary to what might be expected on the basis of previous findings, Yale subjects did not discriminate significantly between the occupations in Categories 9 and 10, even though these were near their presumed anchors.

DISCUSSION

The assumption underlying the present study is that an anchor affects the metric of the subjective reference scale that subjects use in locating stimuli along a judgment dimension. The anchor is seen as exerting an assimilative pull on the reference scale that leads to a compression of the scale. The compression of the scale around the anchor increases the apparent distance between the anchor and new stimulus inputs that differ from the anchor and are referred to the scale. On the response level, this change is manifested as a displacement of the judgments of nonanchor stimuli away from the value assigned to the anchor. A fuller discussion of this view of anchor effects, with some supporting data, may be found in Perloe (1961).

Although the present results support the hypothesis drawn from the approach just described, and are consistent with previous findings related to this approach, another interpretation of the data is possible. This explanation holds that the anchor has its effect by changing the subject's selection of the specific subjective reference scale to be used in judging a set of stimuli rather than by changing the metric of a single scale (McGarvey, 1943; Kelley et al., 1955). According to this view, anchors often determine which aspect of a set of complex stimuli will be salient and therefore which subjective reference scale will be selected for making judgments.

In the present case, the contrast between the extremely desirable statuses held or aspired to by the Yale subjects and the relatively low desirability of the majority of middle and low status occupations they were asked to judge may have led them to make their judgments in terms of their personal view of the desirability of the occupations rather than in terms of their "objective" prestige. If subjects did judge the occupations in this way, it is not surprising that they should find middle and lower class occupations less desirable than the population. These occupations fall much farther below Yale subjects' occupational aspirations than they do below the aspirations of the population at large.

It is interesting to note, in this connection, that the original National Opinion Research Center (1953) study, which required respondents to give "their own personal opinion of the general standing" of occupations, found higher ratings given to most occupations by unskilled laborers, domestic, and service workers than by persons in more prestigeful occupations. This may have been due to the complement of the process suggested for the high status Yale subjects by the alternative explanation. On the other hand, it may also have been due to a metric distortion.

The ambiguity of the quality being judged in the present experiment and in all experiments that ask subjects to locate meaningful stimuli along valued dimensions lends weight to the alternative explanation. It is of course possible that both scale distortion and substitution are involved in these anchor effects. The final evaluation of the importance of the two processes waits upon the application of more adequate techniques for measuring the metric and semantic properties of subjective reference scales.

SUMMARY

A test was made of the hypothesis that high status subjects would displace their judgments of moderate and low prestige occupations toward the low end of the prestige dimension. A comparison of the judgments of Yale students and previously published judgments of a nationwide sample supported the hypothesis. The results were attributed to the high status (Yale) subjects using their own actual or expected positions on the occupational prestige dimension as internal anchors. Two explanations of the anchor effect were discussed. One focused on the distorting effects that anchors have on the metric properties of subjective reference scales; the other stressed the role played by anchors in determining which of several potential subjective reference scales would be used in judging a set of stimuli.

REFERENCES

GUILFORD, J. P. *Psychometric methods.* (2nd ed.) New York: McGraw-Hill, 1954.

HELSON, H. Adaptation level as a basis for a quantita-

[5] The Mann-Whitney U test was employed.

tive theory of frames of reference. *Psychol. Rev.*, 1948, **55,** 297–313.

HINCKLEY, E. D., & RETHLINGSHAFER, D. Value judgments of heights of men by college students. *J. Psychol.*, 1951, **31,** 257–262.

HOVLAND, C. I., HARVEY, O. J., & SHERIF, M. Assimilation and contrast effects in reactions to communication and attitude change. *J. abnorm. soc. Psychol.*, 1957, **55,** 244–252.

HOVLAND, C. I., & SHERIF, M. Judgmental phenomena and scales of attitude measurement: Item displacement in Thurstone scales. *J. abnorm. soc. Psychol.*, 1952, **47,** 822–832.

HUNT, W. A. Anchoring effects in judgment. *Amer. J. Psychol.*, 1941, **54,** 395–403.

KELLEY, H. H., HOVLAND, C. I., SCHWARTZ, M., & ABELSON, R. P. The influence of judges' attitudes in three methods of attitude scaling. *J. soc. Psychol.*, 1955, **42,** 147–158.

MCGARVEY, H. R. Anchoring effects in the absolute judgment of verbal materials. *Arch. Psychol. N. Y.*, 1943, **39,** No. 281.

MANIS, M. The interpretation of opinion statements as a function of recipient attitude. *J. abnorm. soc. Psychol.*, 1960, **60,** 340–344.

NATIONAL OPINION RESEARCH CENTER. Jobs and occupations: A popular evaluation. In R. Bendix & S. M. Lipset (Eds.), *Class, status and power: A reader in social stratification.* Glencoe, Ill.: Free Press, 1953. Pp. 411–426.

PERLOE, S. I. Assimilation as a consequence of categorization. *J. Pers.*, 1961, **29,** 148–166.

SEGALL, M. H. The effect of attitude and experience on judgments of controversial statements. *J. abnorm. soc. Psychol.*, 1959, **58,** 61–68.

SHERIF, M., & HOVLAND, C. I. Judgmental phenomena and scales of attitude measurement: Placement of items with individual choice of number of categories. *J. abnorm. soc. Psychol.*, 1953, **48,** 135–141.

TRESSELT, MARGARET E. The effect of the experience of contrasted groups upon the formation of a new scale of judgment. *J. soc. Psychol.*, 1948, **27,** 209–216.

VOLKMANN, J. Scales of judgment and their implications for social psychology. In J. H. Rohrer & M. Sherif (Eds.), *Social psychology at the crossroads.* New York: Harper, 1951. Pp. 273–294.

WALKER, HELEN M., & LEV, J. *Statistical inference.* New York: Holt, 1953.

WEISS, W. The effects on opinions of a change in scale judgments. *J. abnorm. soc. Psychol.*, 1959, **58,** 329–334.

(Received December 27, 1960)

Chestnut Lodge Symposium

PAPERS PRESENTED ON THE 50th ANNIVERSARY, 1910–1960

Contents

DEXTER M. BULLARD: Introduction

JOHN C. WHITEHORN: Alienation and Leadership

DAVID McK. RIOCH: The Sense and the Noise

ALFRED H. STANTON: Milieu Therapy and the Development of Insight

DON D. JACKSON and JOHN H. WEAKLAND: Conjoint Family Therapy

ROBERT A. COHEN and MABEL BLAKE COHEN: Research in Psychotherapy

CLARENCE G. SCHULZ: Case Report of an Obsessional Patient

OTTO ALLEN WILL, JR.: Paranoid Development and the Concept of Self: Psychotherapeutic Intervention

HAROLD F. SEARLES: Sexual Processes in Schizophrenia

DONALD L. BURNHAM: Identity Definition and Role Demand in the Hospital Careers of Schizophrenic Patients

$3.50

William Alanson White Psychiatric Foundation

1610 New Hampshire Ave., N.W., Washington 9, D.C.

An interdisciplinary journal for behavioral and social scientists who wish to keep abreast of research concerned with the problems of peace and international relations

The Journal of

conflict

RESOLUTION

A quarterly for research related to war and peace

Annual subscription: $5.50
per copy: $2.00

Published by The Center for Research on Conflict Resolution
820 East Washington, University of Michigan, Ann Arbor, Michigan

Neo-Freudian Social Philosophy

MARTIN BIRNBACH

The implications of psychoanalytic theory for contemporary social and political problems are outlined in this analysis of the writings of Franz Alexander, Erich Fromm, Karen Horney, Abram Kardiner, and Harry Stack Sullivan. In addition, there is an exposition of Harold D. Lasswell's important work on the application of the psychoanalytic viewpoint to politics. The main study is preceded by an explication of the principal elements of the Freudian system and their philosophical import, a description of the break with Freud's concepts, and a statement of Neo-Freudian positions on psychoanalytic theory. Extensive Bibliography and biographical sketches. $6.00

Perceval's Narrative

A PATIENT'S ACCOUNT OF HIS PSYCHOSIS, 1830–1833

JOHN PERCEVAL

Edited and with an Introduction by GREGORY BATESON

John Perceval wrote this autobiographical account of his three years of schizophrenic illness long before there were any psychiatric theories regarding schizophrenia. His story is therefore important from a scientific point of view because it is uninfluenced by Freudian and other ideas which have subsequently been proposed. Perceval tells of his illness and recovery with great vigor and insight, giving a surprisingly modern view of schizophrenia. He makes an attempt to explain why he became mentally ill and why his particular symptoms developed, and recounts a number of episodes which contributed to his recovery. $6.75

Order from your bookstore, please

STANFORD UNIVERSITY PRESS

NEW EDITION!!

"The best in tests....."

Could you use the new, 1960 edition of this well-known test for brain injury and perceptual-motor coordination, widely used in clinical practice and in research since 1945, with new norms based on large clinical and control groups?

Graham-Kendall

MEMORY FOR DESIGNS TEST

Then order Tester's Set consisting of durable plates and 1960 manual ($8.50), or just the 1960 manual ($2.50) if you already have plates.

•

Obtain further information by writing

PSYCHOLOGICAL TEST
SPECIALISTS
Box 1441
Missoula, Montana

"The best in tests....."

NEW EDITION!!

1553